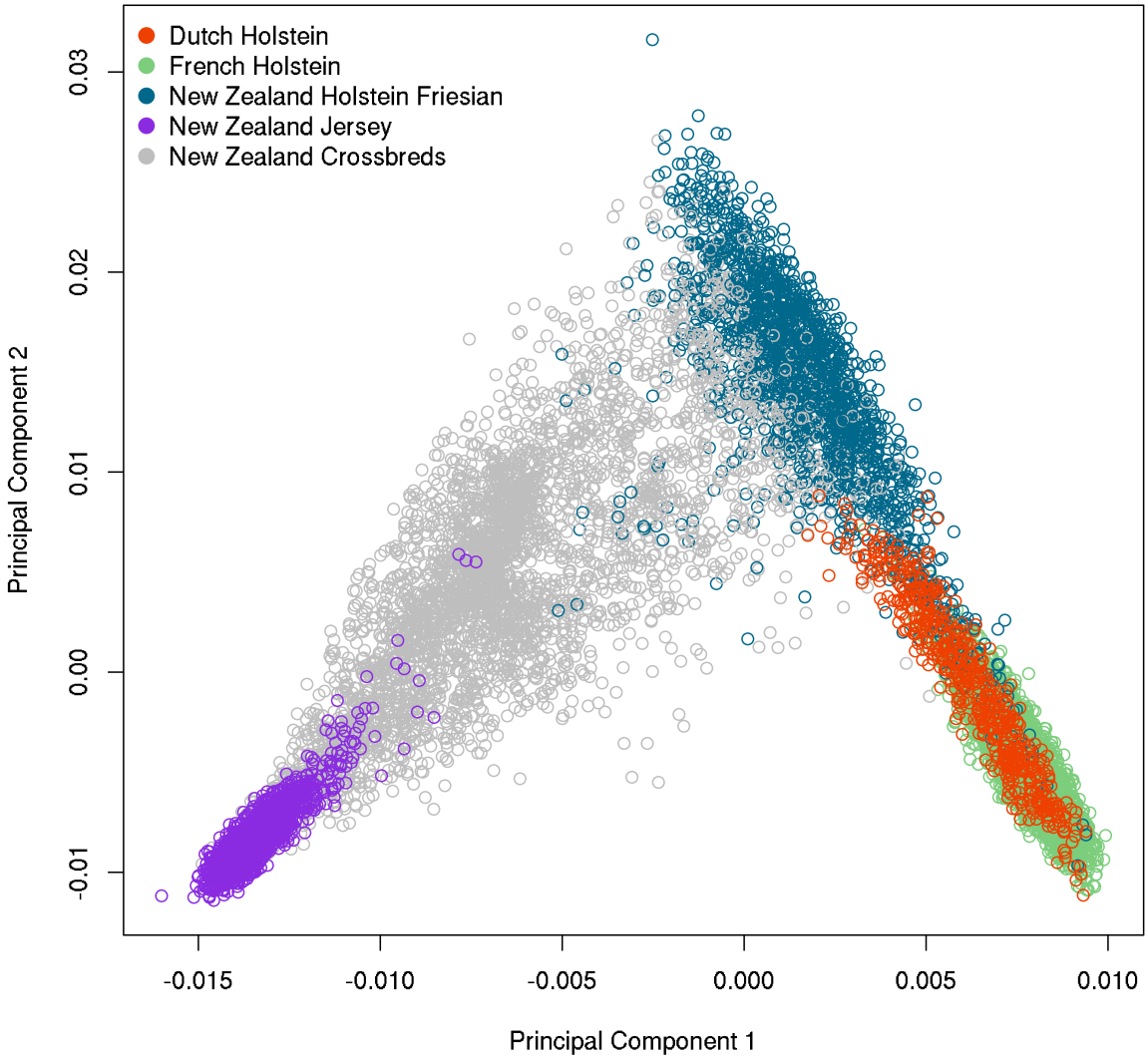
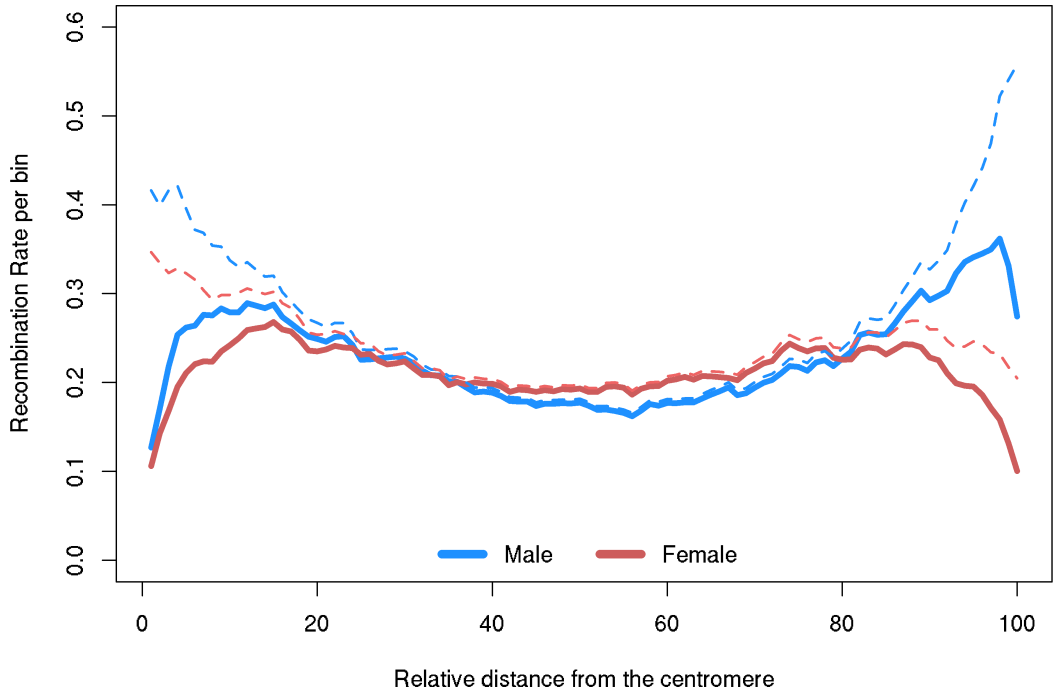
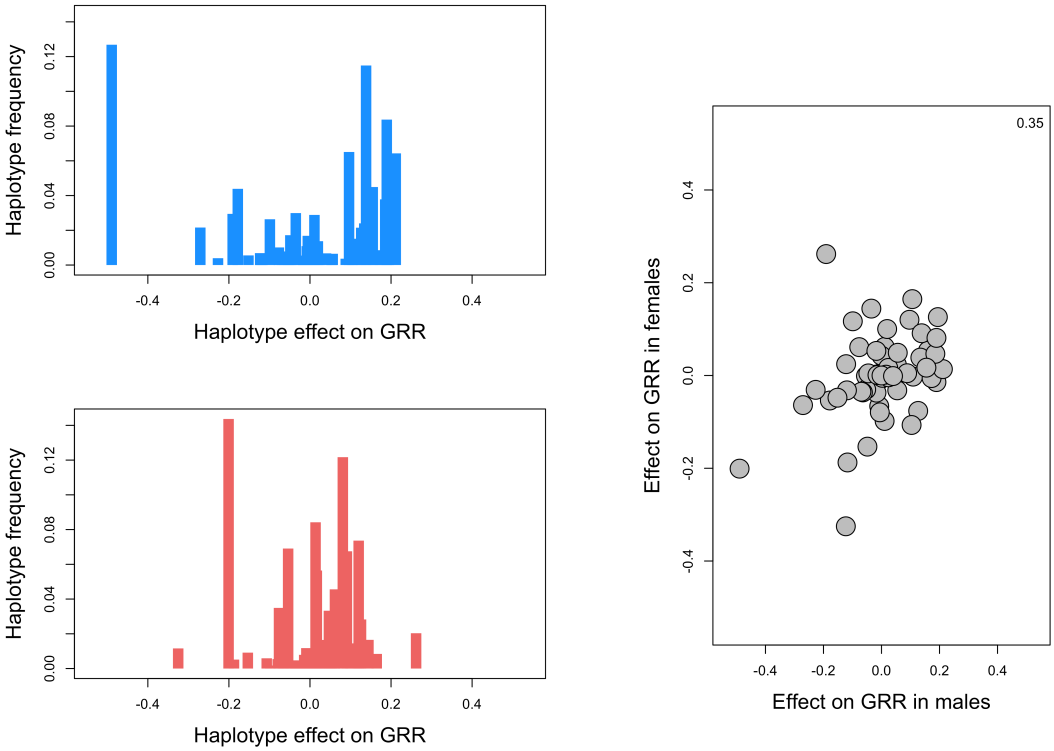


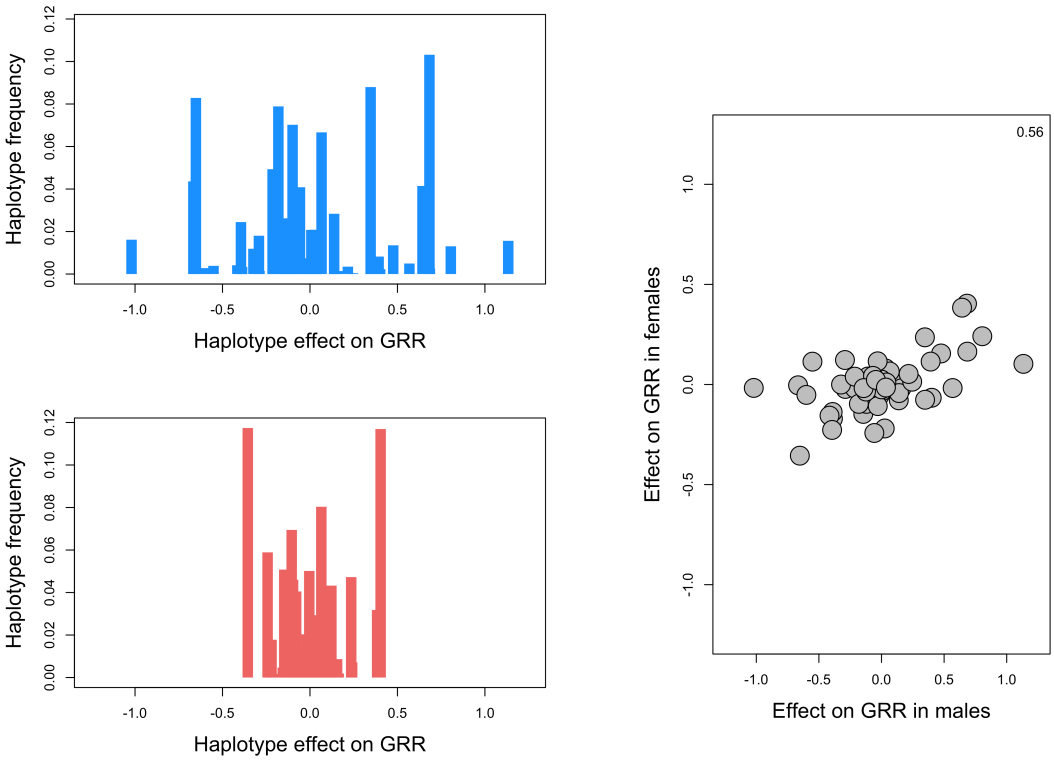
Supplemental Figure S1. Principal Component Analysis of genotyped proband. Individuals are plotted according to their coordinates on the two first principal components. The analysis was done with the GCTA software package (Yang et al. 2011) and using the 30,127 SNPs selected to study global recombination rate.



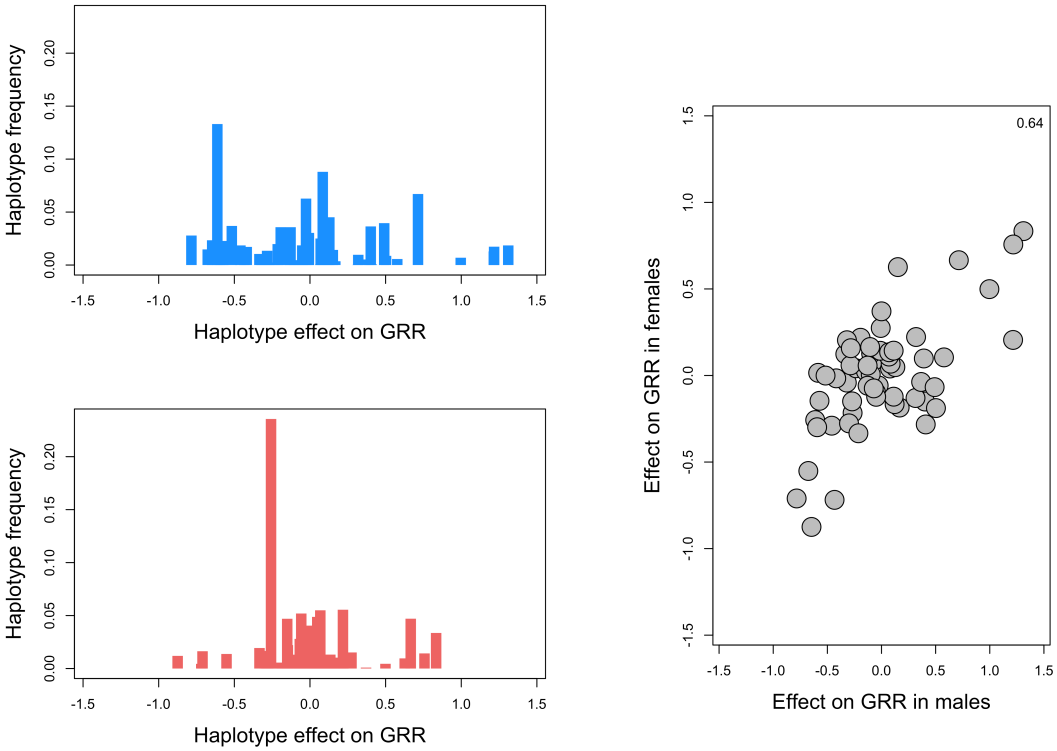
Supplemental Figure S2. Distribution of CO events along the chromosomes in females (red) and males (blue). After removal of 2 Mb on both extremities, chromosomes were subdivided into 100 equally sized bins and number of CO events observed in each bin summed across the 29 autosomes. Resulting numbers were then divided by the number of analyzed oocytes and sperm cells. As a consequence, the sum of the corresponding numbers across bins amounts to the female and male map lengths, respectively. The full lines correspond to the actual data. The dotted lines correspond to the actual data divided by an estimate of the probability to detect a CO event in the corresponding bin if it exists, i.e. the local informativity. We considered that a CO must be flanked by two informative markers (heterozygous in the parent and phased in both parent and offspring) on both sides to be detected with our method. We estimated the informativity at a given position as the proportion of meiosis fulfilling that condition.



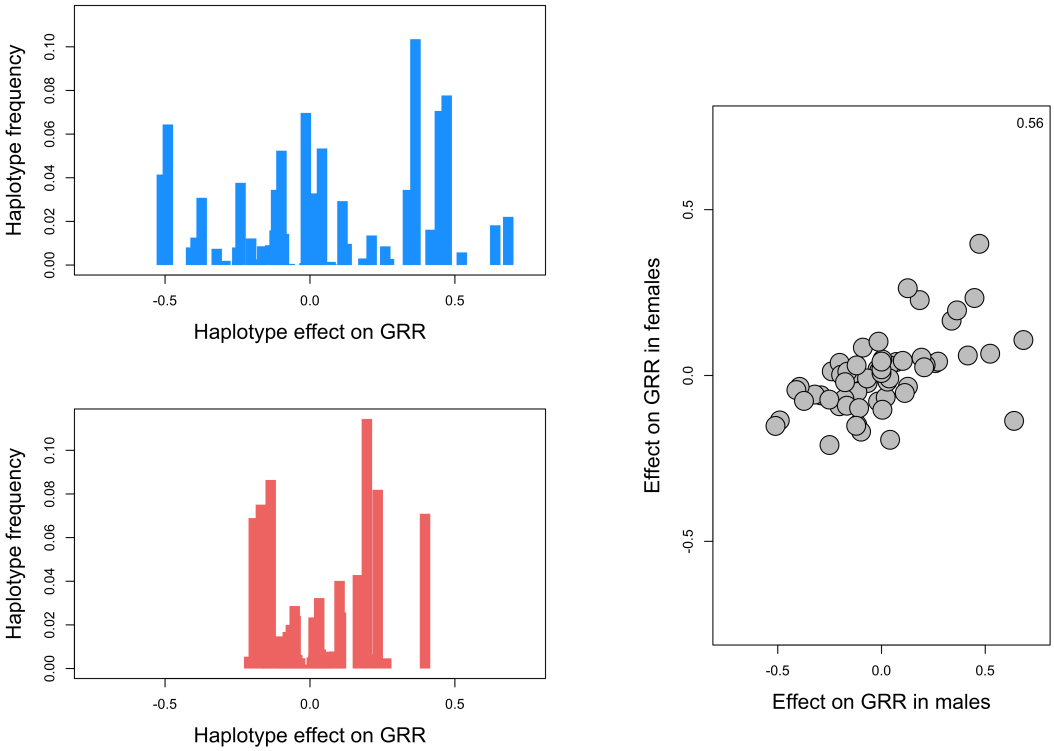
Supplemental Figure S3. Ancestral Haplotypes effects on male and female GRR - BTA3 QTL. Distribution of haplotype effects estimated on BTA3 (at position 52.4 Mb). The blue histogram represents the frequency of the 60 ancestral haplotypes groups (y-axis) and their estimated effects on male GRR (x-axis). The red histogram represents estimated effects for female GRR. The right plot compares for the 60 ancestral haplotypes their solutions in males (x-axis) and in females (y-axis). The corresponding correlation is indicated in the right corner.



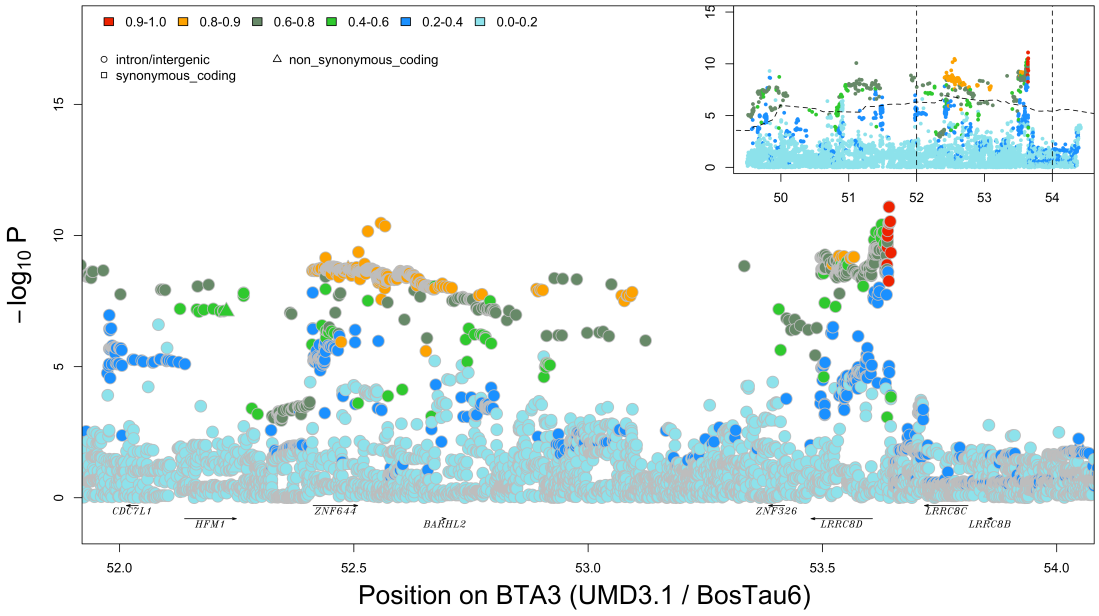
Supplemental Figure S4. Ancestral Haplotypes effects on male and female GRR - BTA6 QTL. Distribution of haplotype effects estimated on BTA6 (at position 122,4 Mb (corrected map)). The blue histogram represents the frequency of the 60 ancestral haplotypes groups (y-axis) and their estimated effects on male GRR (x-axis). The red histogram represents estimated effects for female GRR. The right plot compares for the 60 ancestral haplotypes their solutions in males (x-axis) and in females (y-axis). The corresponding correlation is indicated in the right corner.



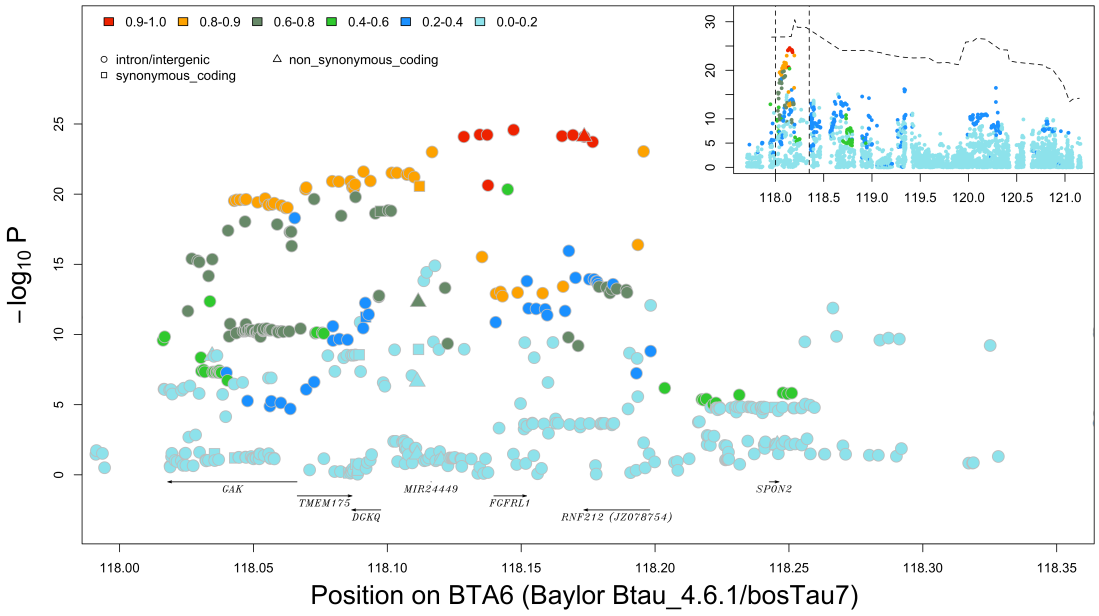
Supplemental Figure S5. Ancestral Haplotypes effects on male and female GRR - BTA10 proximal QTL. Distribution of haplotype effects estimated on BTA10 (at position 21.20 Mb). The blue histogram represents the frequency of the 60 ancestral haplotypes groups (y-axis) and their estimated effects on male GRR (x-axis). The red histogram represents estimated effects for female GRR. The right plot compares for the 60 ancestral haplotypes their solutions in males (x-axis) and in females (y-axis). The corresponding correlation is indicated in the right corner.



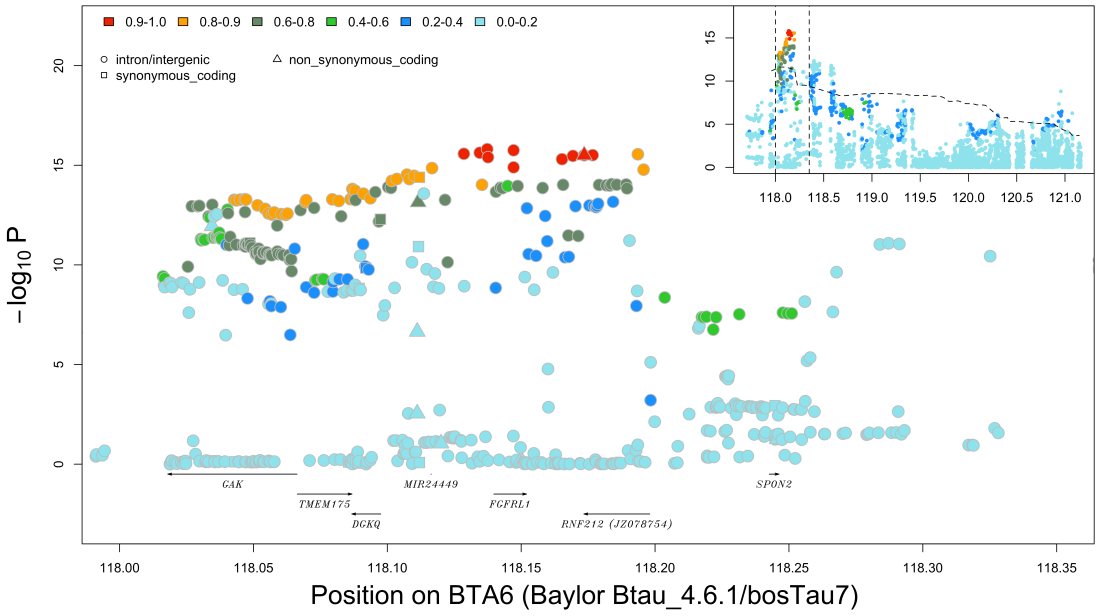
Supplemental Figure S6. Ancestral Haplotypes effects on male and female GRR - BTA10 distal QTL. Distribution of haplotype effects estimated on BTA10 (at position 86.5 Mb). The blue histogram represents the frequency of the 60 ancestral haplotypes groups (y-axis) and their estimated effects on male GRR (x-axis). The red histogram represents estimated effects for female GRR. The right plot compares for the 60 ancestral haplotypes their solutions in males (x-axis) and in females (y-axis). The corresponding correlation is indicated in the right corner.



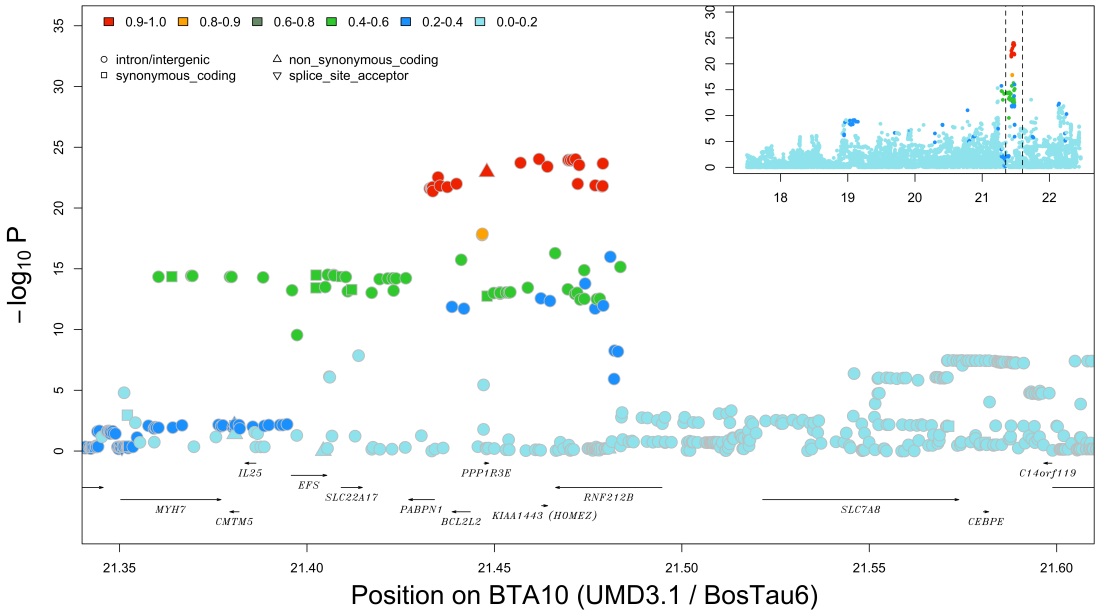
Supplemental Figure S7. Sequence-based association analysis in males for the BTA3 QTL (52.40 Mb) associated with *HFM1*. The variants in red define "LD-based set of candidate variants" assumed to encompass the causative variants. The dashed line in the inset represents the significance of the haplotype-based association signal, while the dotted vertical lines define the boundaries of the region that are zoomed in the main graph.



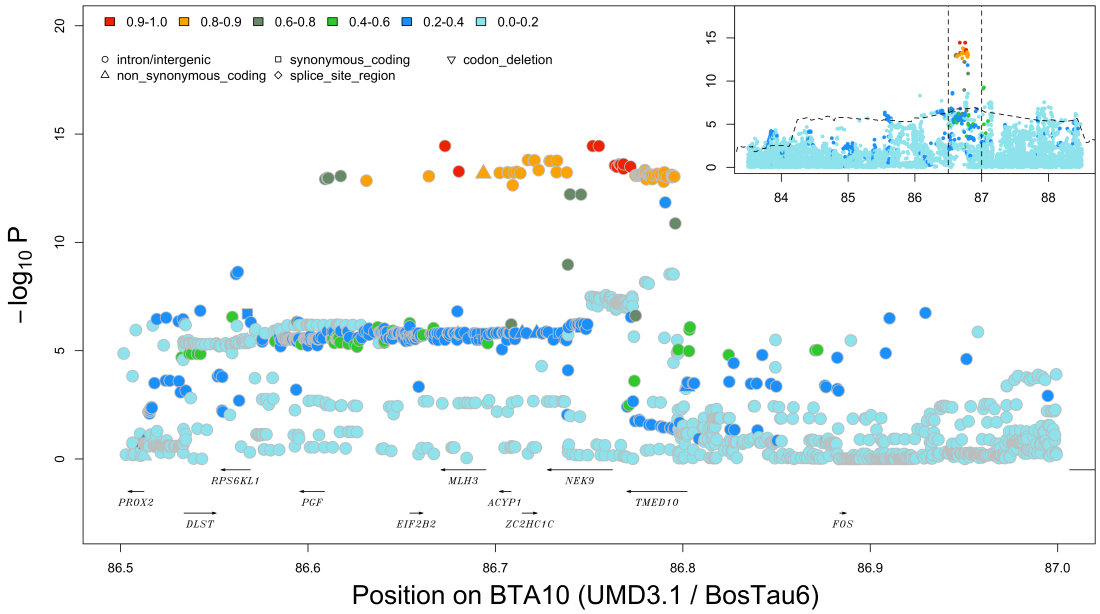
Supplemental Figure S8. Sequence-based association analysis in males for the primary signal of BTA6 QTL associated with *RNF212*. The variants in red define "LD-based set of candidate variants" assumed to encompass the causative variants. The dashed line in the inset represents the significance of the haplotype-based association signal, while the dotted vertical lines define the boundaries of the region that are zoomed in the main graph.



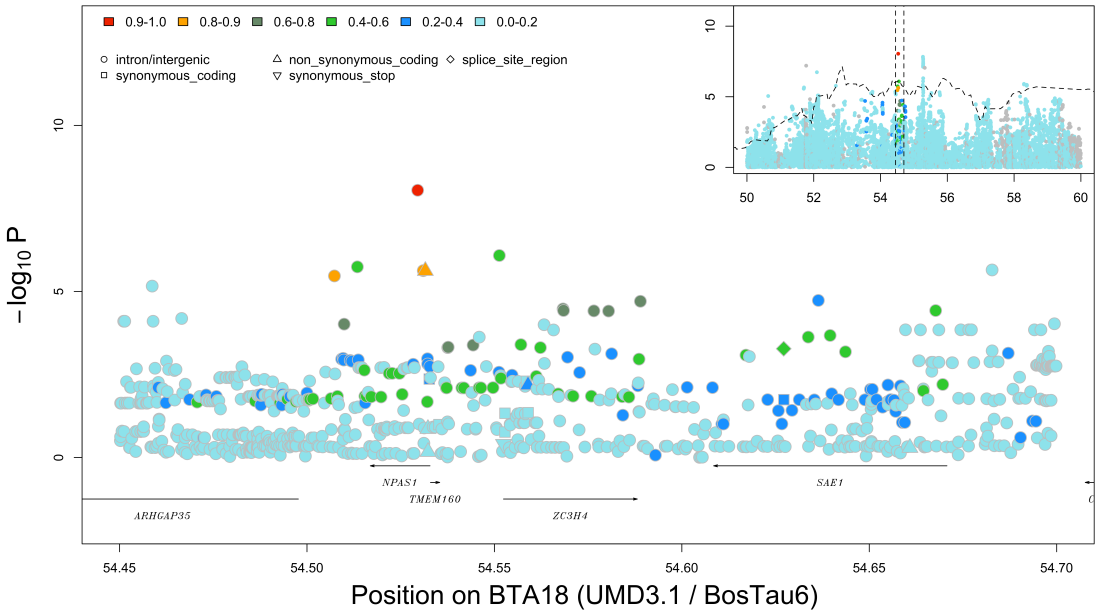
Supplemental Figure S9. Sequence-based association analysis in females for the primary signal of BTA6 QTL associated with *RNF212*. The variants in red define "LD-based set of candidate variants" assumed to encompass the causative variants. The dashed line in the inset represents the significance of the haplotype-based association signal, while the dotted vertical lines define the boundaries of the region that are zoomed in the main graph.



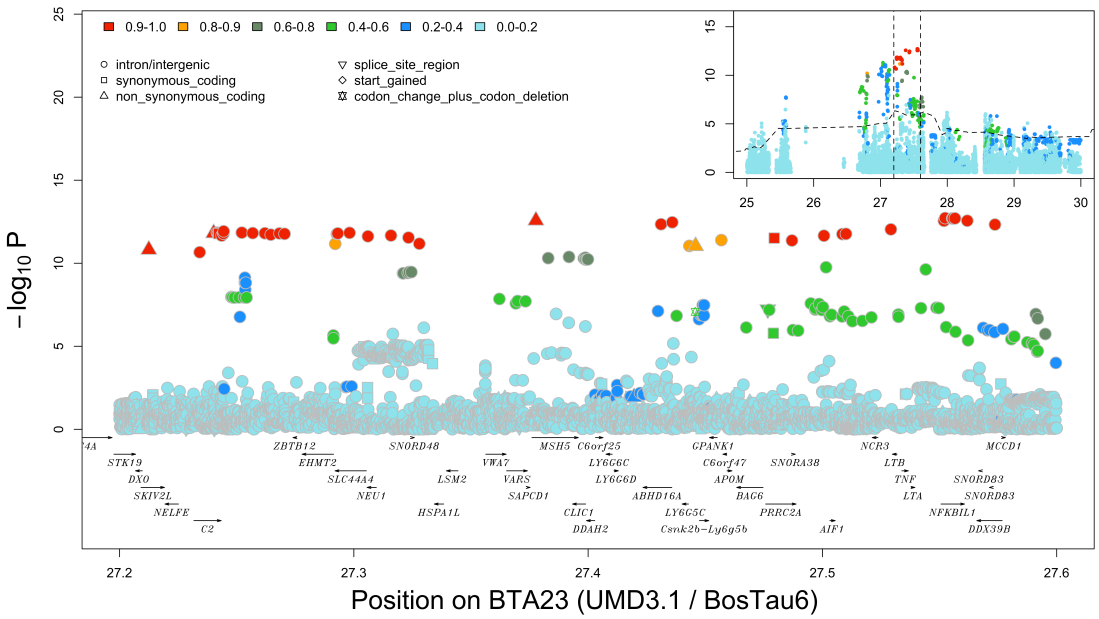
Supplemental Figure S10. Sequence-based association analysis in females for the primary signal of BTA10 QTL (21.20 Mb) associated with *RNF212B*. The variants in red define "LD-based set of candidate variants" assumed to encompass the causative variants. The dashed line in the inset represents the significance of the haplotype-based association signal, while the dotted vertical lines define the boundaries of the region that are zoomed in the main graph.



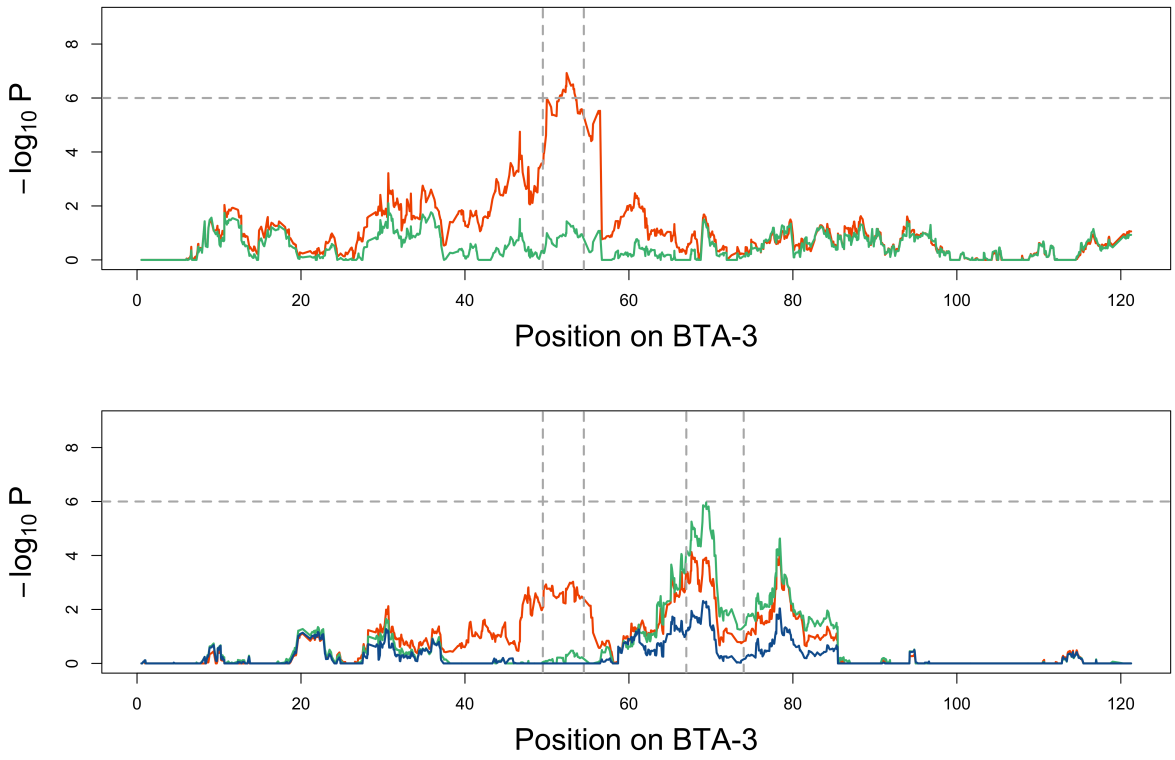
Supplemental Figure S11. Sequence-based association analysis in females for the BTA10 QTL (86.51 Mb) associated with *MLH3*. The variants in red define "LD-based set of candidate variants" assumed to encompass the causative variants. The dashed line in the inset represents the significance of the haplotype-based association signal, while the dotted vertical lines define the boundaries of the region that are zoomed in the main graph.



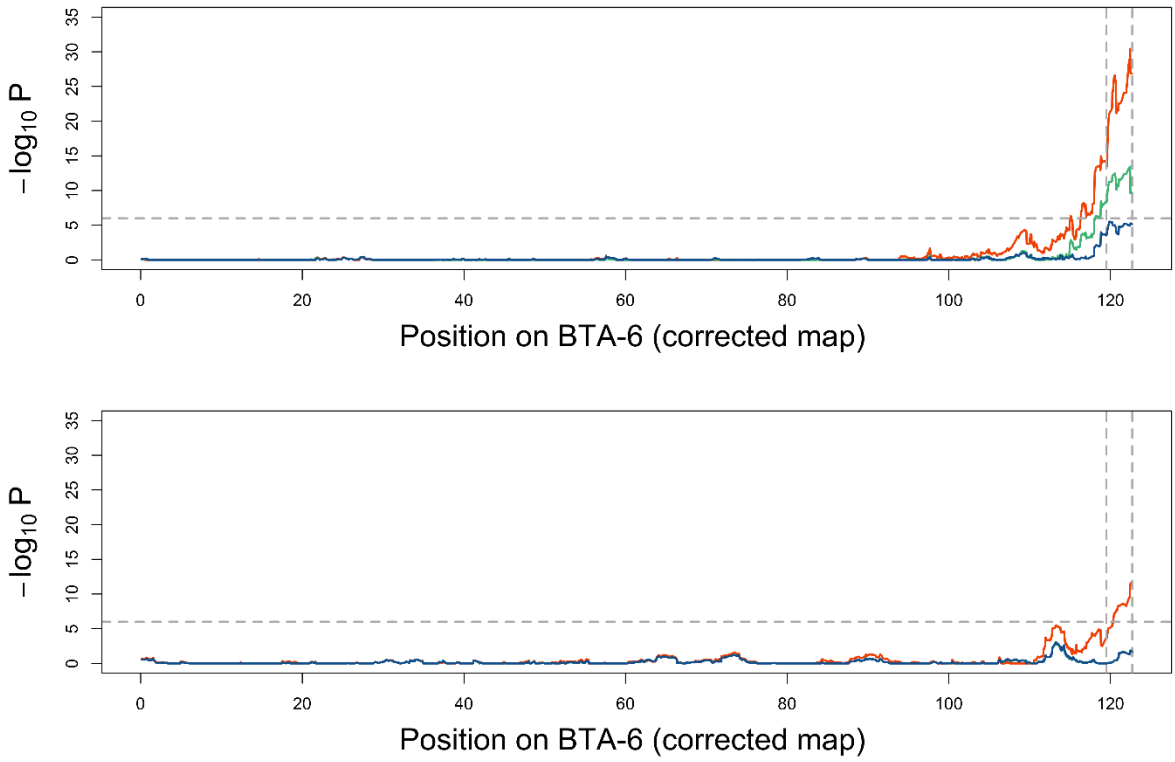
Supplemental Figure S12. Sequence-based association analysis in males for the BTA18 QTL (52.9 Mb). The variants in red define "LD-based set of candidate variants" assumed to encompass the causative variants. The dashed line in the inset represents the significance of the haplotype-based association signal, while the dotted vertical lines define the boundaries of the region that are zoomed in the main graph.



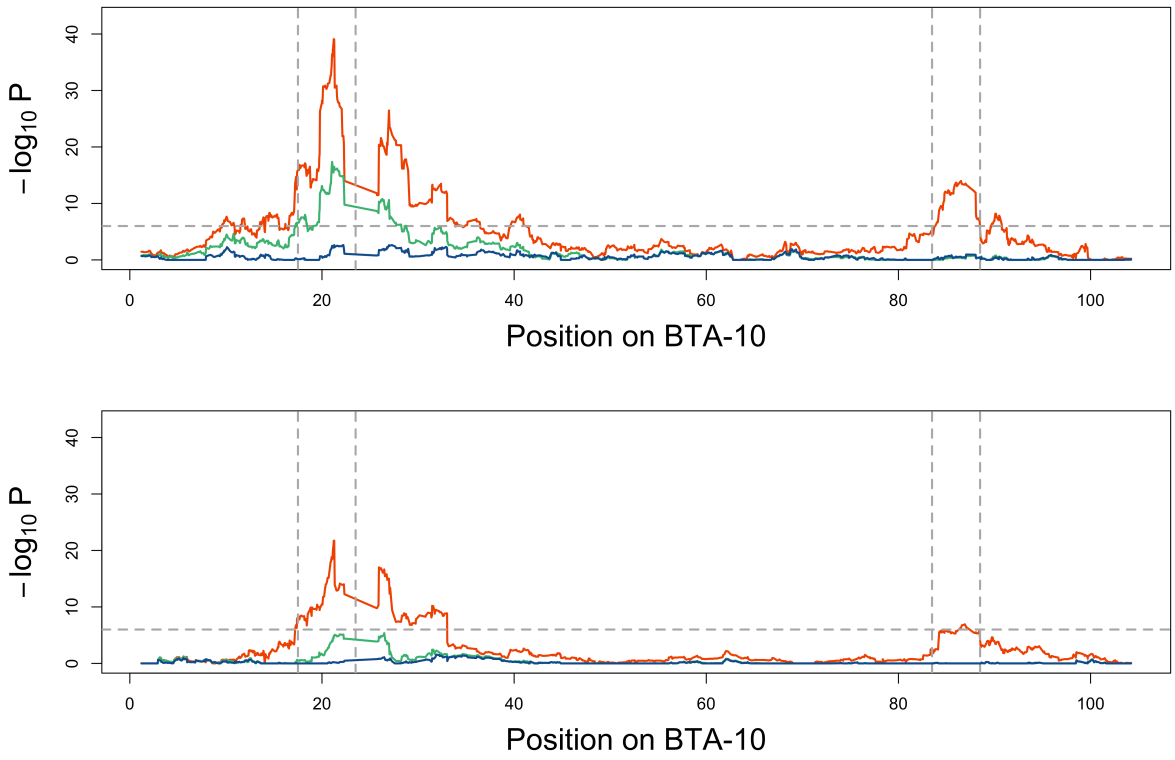
Supplemental Figure S13. Sequence-based association analysis in males for the BTA23 QTL associated with *MSH5*. The variants in red define "LD-based set of candidate variants" assumed to encompass the causative variants. The dashed line in the inset represents the significance of the haplotype-based association signal, while the dotted vertical lines define the boundaries of the region that are zoomed in the main graph.



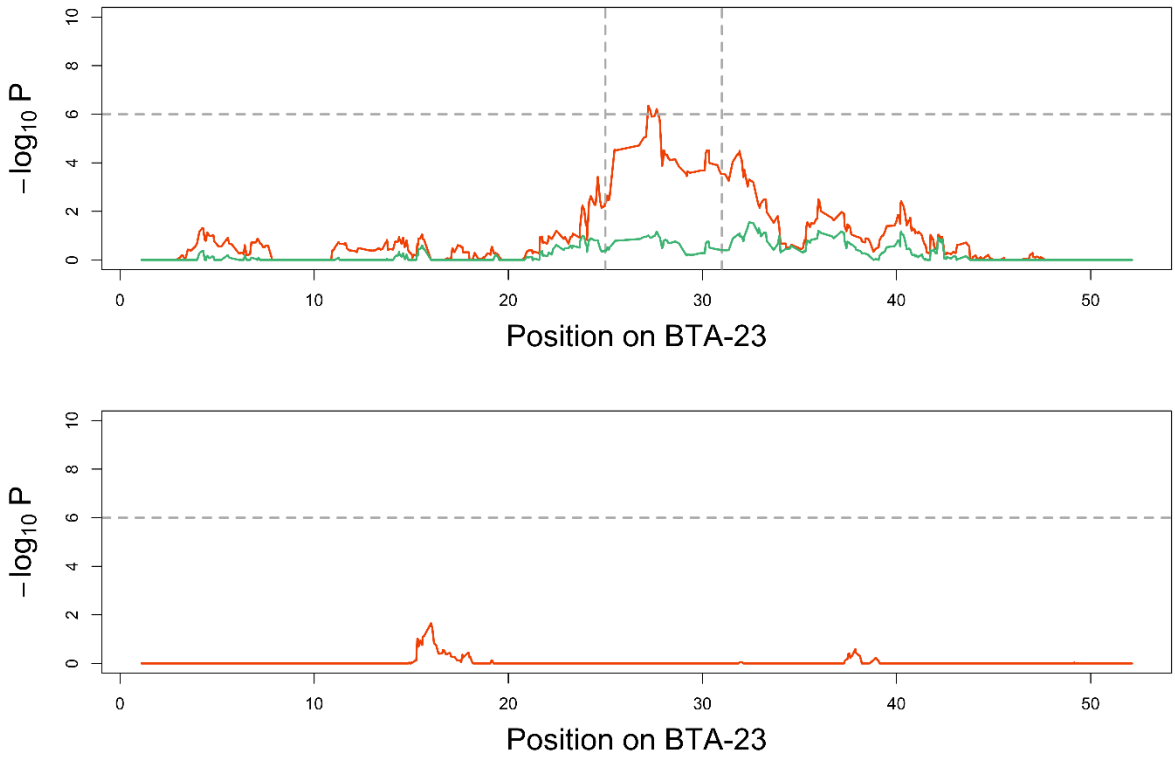
Supplemental Figure S14. Haplotype-based association studies on BTA3. Male (upper panel) and female (lower panel) haplotype-based association studies obtained 1) with fitting the 60 ancestral haplotypes effects along BTA3 (red), 2) adding to model 1 the *HFM1 S1189L* variant (green) and 3) adding to model 2 the intronic *MSH4* variant rs210318688 (blue).



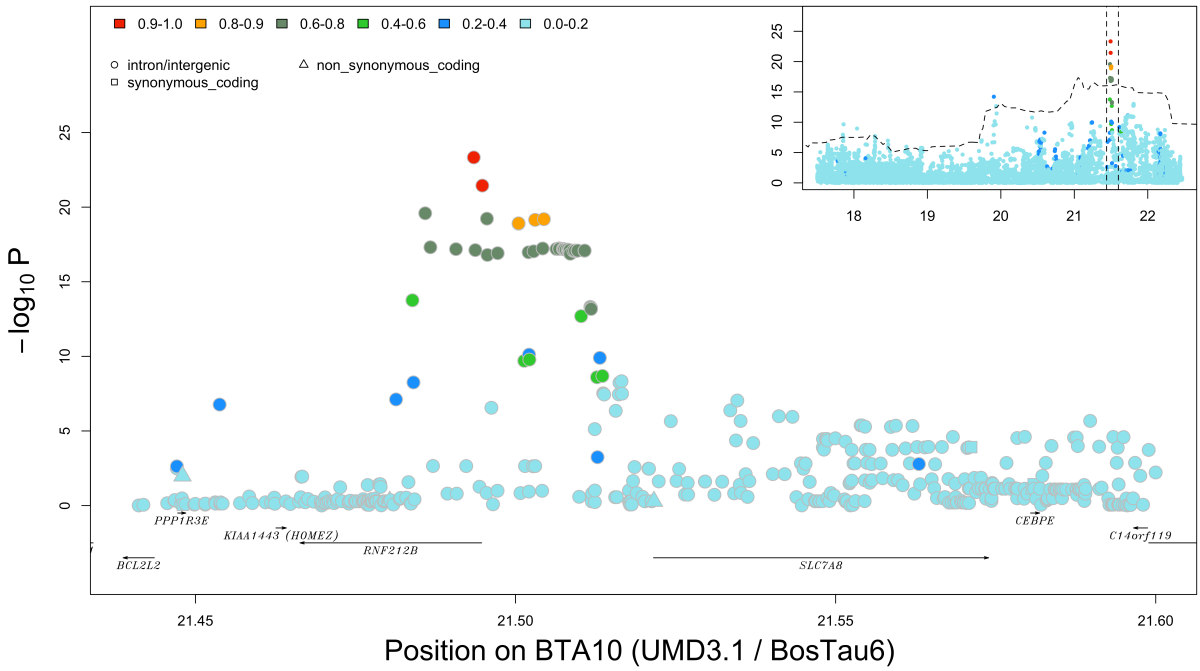
Supplemental Figure S15. Haplotype-based association studies on BTA6. Male (upper panel) and female (lower panel) haplotype-based association studies obtained 1) with fitting the 60 ancestral haplotypes effects along BTA6 (red), 2) adding to model 1 the *RNF212 P259S* variant (green) and 3) adding to model 2 the *RNF212* A77T variant (blue). A corrected map of BTA6 was used as described in Supplementary Note 4.



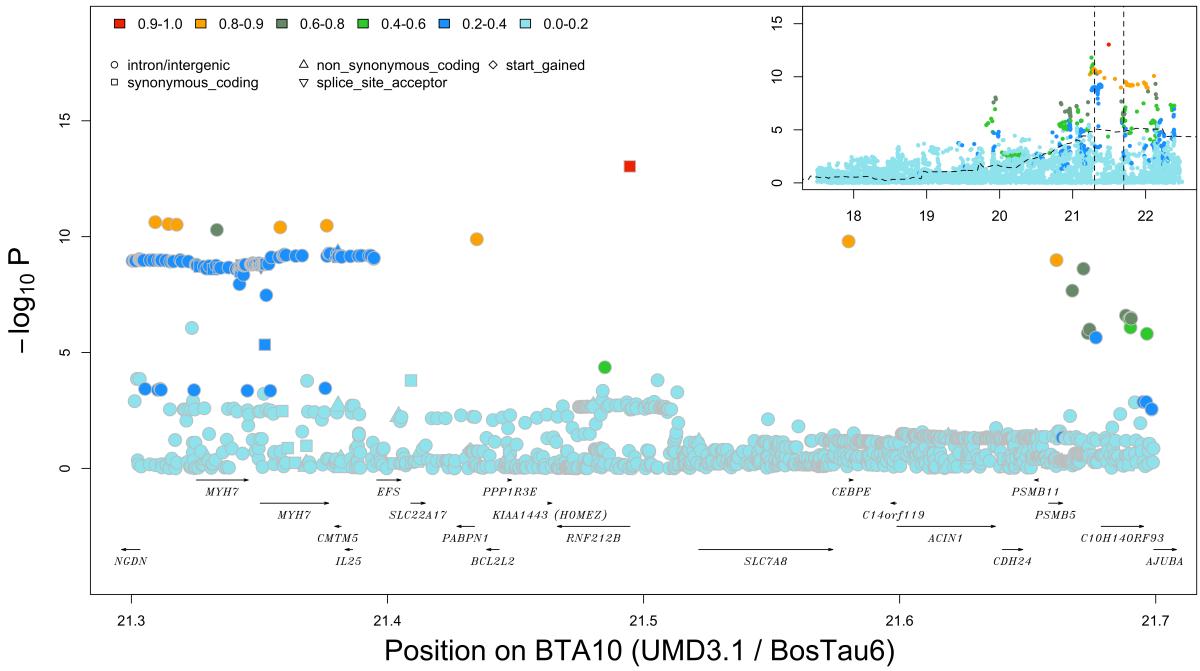
Supplemental Figure S16. Haplotype-based association studies on BTA10. Male (upper panel) and female (lower panel) haplotype-based association studies obtained 1) with fitting the 60 ancestral haplotypes effects along BTA10 (red), 2) adding to model 1 the lead variant associated to *RNF212B* (rs381356614) and the *MLH3* N408S variant (green) and 3) adding to model 2 secondary variants rs207682689 and rs437013002 associated to RNF212B (blue).



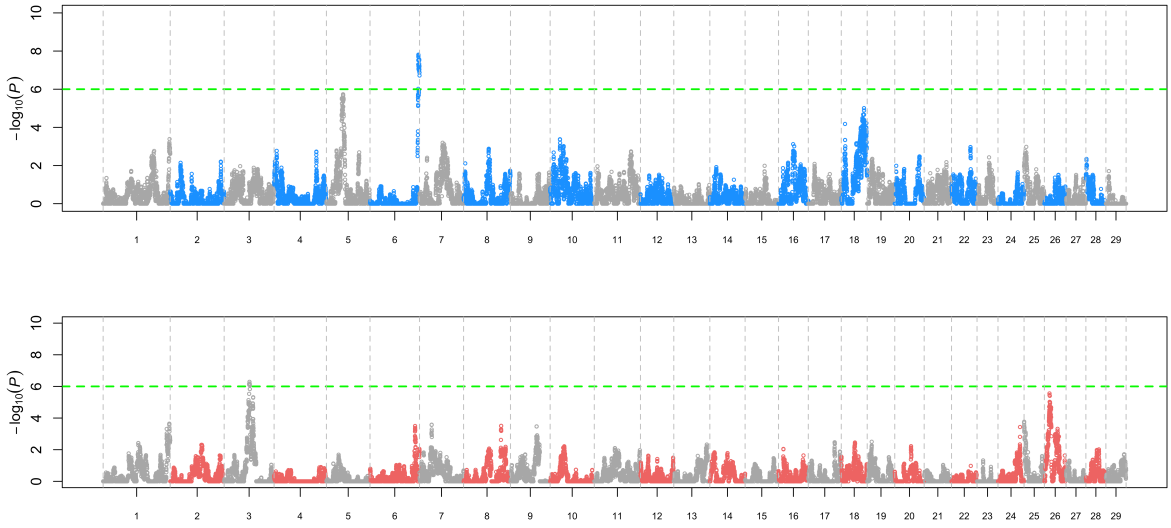
Supplemental Figure S17. Haplotype-based association studies on BTA23. Male (upper panel) and female (lower panel) haplotype-based association studies obtained 1) with fitting the 60 ancestral haplotypes effects along BTA23 (red) and 2) adding to model 1 the *MSH5 R631Q* variant (green).



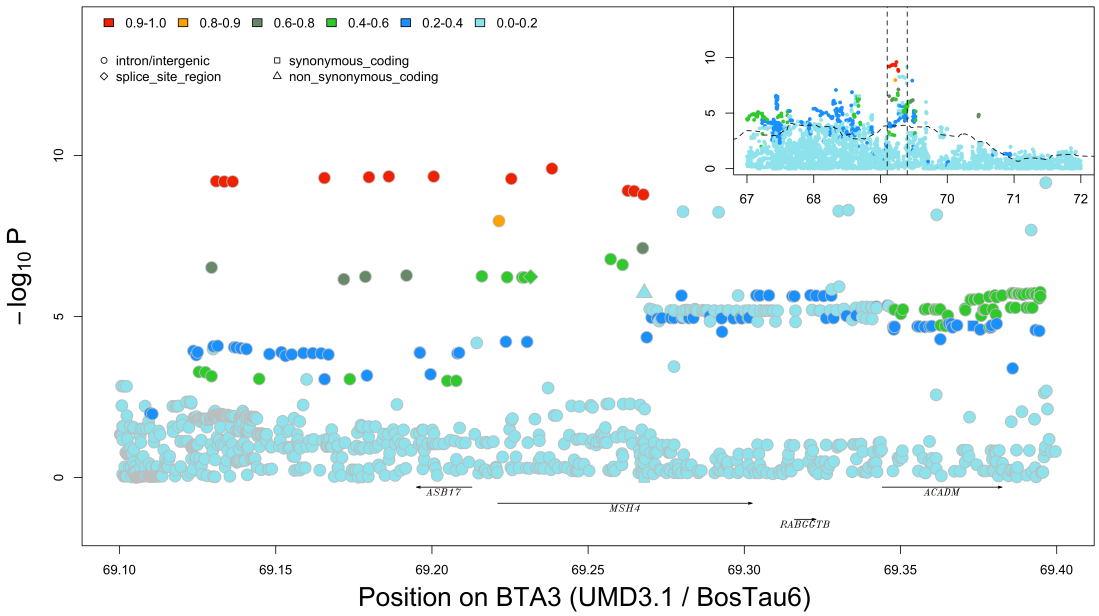
Supplemental Figure S18. Secondary sequence-based association analysis in males for the BTA10 QTL (21.20 Mb) associated with *RNF212B.* The model includes the *RNF212* P259S, the *MLH3* N408S, the *HFM1* S1189L, the *MSH5* R631Q variants and the *RNF212B* associated variant rs381356614 as covariates. The variants in red define "LD-based set of candidate variants" assumed to encompass the causative variants. The dashed line in the inset represents the significance of the haplotype-based association signal, while the dotted vertical lines define the boundaries of the region that are zoomed in the main graph.



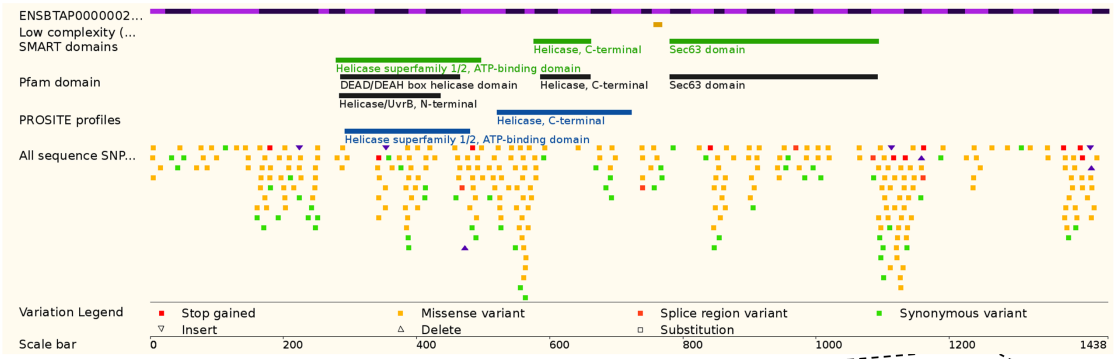
Supplementl Figure S19. Secondary sequence-based association analysis in females for the BTA10 QTL (21.20 Mb) associated with *RNF212B.* The model includes the *RNF212* P259S, the *MLH3* N408S, the *HFM1* S1189L, the *MSH5* R631Q variants and the *RNF212B* associated variant rs381356614 as covariates. The variants in red define "LD-based set of candidate variants" assumed to encompass the causative variants. The dashed line in the inset represents the significance of the haplotype-based association signal, while the dotted vertical lines define the boundaries of the region that are zoomed in the main graph.



Supplemental Figure S20. Second Manhattan plot for male and female GRR in cattle. The significance (log(1/p)) of the haplotype-based association is reported along the genome for male (blue) and female (red) GRR. The association is performed with a model including the *RNF212* P259S, *RNF212* A77T (in males), the *MLH3* N408S, the *HFM1* S1189L, the *MSH5* R631Q (in males), the *RNF212B* associated variants (rs381356614, rs207682689, rs437013002) and rs135941180 (in males) as covariates.



Supplemental Figure S21. Sequence-based association analysis in females for the BTA3 QTL (69.42 Mb). The model includes the *RNF212* P259S, the *MLH3* N408S, the *HFM1* S1189L and the *RNF212B* associated variants (rs381356614 and rs437013002) as covariates. The variants in red define "LD-based set of candidate variants" assumed to encompass the causative variants. The dashed line in the inset represents the significance of the haplotype-based association signal, while the dotted vertical lines define the boundaries of the region that are zoomed in the main graph.



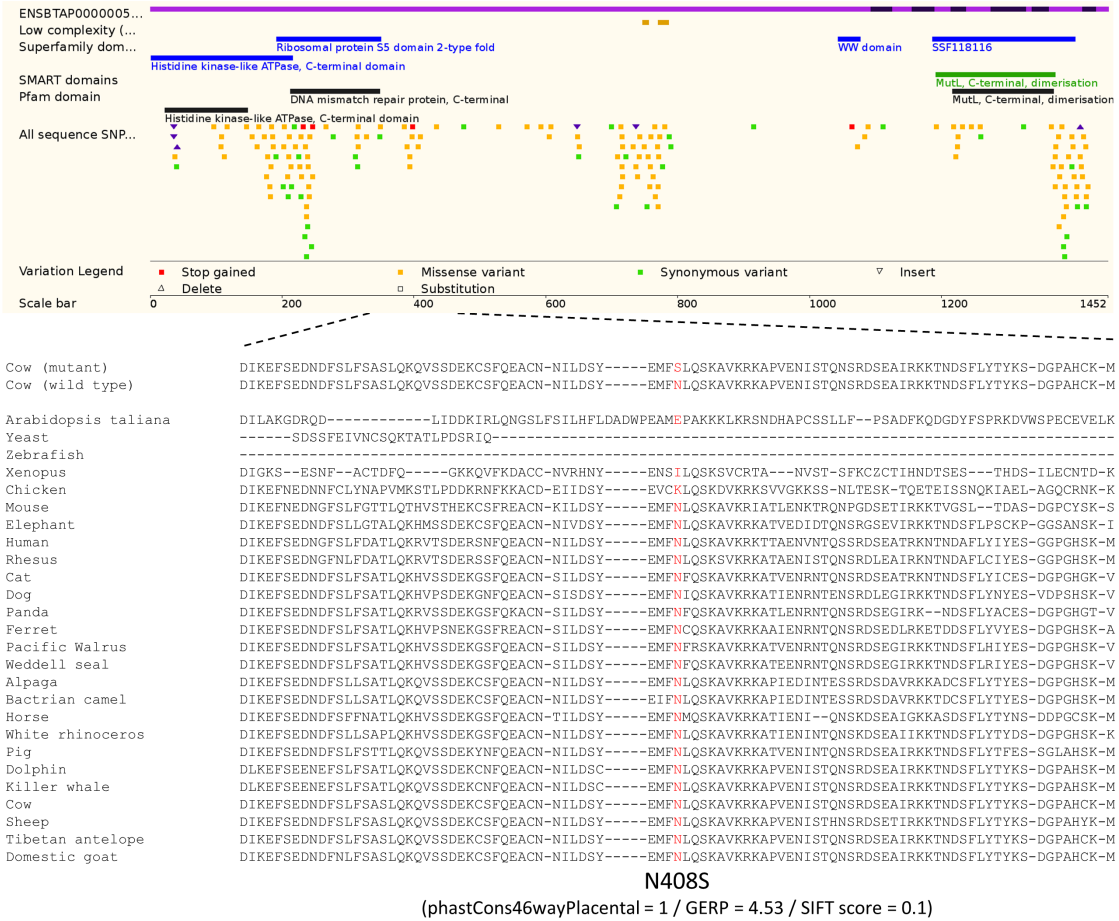

```
Cow (mutant)            PGNRECNHHCKNKHTC-GHDCCKTGVAQKSEGKESTISSYLSDLRNRNAVLSLPPVKRLKMQMNKSPSVDLKDFGFAPKPSLPSIARSE--YLNTPELPIM
Cow (wild type)         PGNRECNHHCKNKHTC-GHDCCKTGVAQKSEGKESTISSYLSDLRNRNAVSSLPPVKRLKMQMNKSPSVDLKDFGFAPKPSLPSIARSE--YLNTPELPIM

Yeast                   NGNYECFHSCKDKTQC-RHLCCKEGIPVKYIKEK-------------GPSSIKPVSKPDQIR----QPLLAKNINTTPHLEKRLNSK----------PKQ
Arabidopsis taliana     IDDD-SEEEKEPYVTMEEDDCVIINEH--------TV---FDHIRE--KAKCFPSLNPLNPTS-------------SPASGKSILKRKSLVENNSPELDPL
Zebrafish               ----------------------------------------------------------------------------------------------------
Xenopus                 ----------------------------------------------------------------------------------------------------
Chicken                 DS------------------------------------------------------------------------------------------------
Rhesus                  PGNRECNHPCKSKHTC-GHNCCKIGVAQKSEIKESTISSYLSDLRNRNAVSSVPTVKRLK--------------------------SE--Y---------
Mouse                   PGNRECHHHCKNKHAC-GHDCCKIGVAQKPEVKESAMSSYLSDLKSRDAVSSLPLAKRLKIQMNKSQNVDLKEFGFTPRPSLSSISRSE--YLNTPELSIL
Elephant                LGNRECNHHCKNKHTC-GHDCCKVGVAQKSGVKEPTISSYLSDLRNRNAVSSLPPVKRLKLQMNKSQSVNLKEFGFTPKSSLPGISRSE--YSNILELPIM
Alpaga                  PGNRECNHHCKNKHTC-GHDCCKTGVAQKSEMKESTISSYISDLRSRDAVSSLPPVKRLKMQMNKSQSVNLKEFGFTPKPSLPSISRSE--YLNTPEVPMM
Bactrian camel          PGNRECNHHCKNKHTC-GHDCCKTGVAQKSEMKGSTISSYISDLRSRDAVSSLPPVKRLKMQMNKSQSVNLKEFGFTPKTSLPSISRSE--YLNTPEVPMM
Pig                     PGHRECNHHCKKKHTC-GHDCCKTGVAQKPEMKESTISSYLSDLRNRNAVSSLPPVKRLKMQMNKSQSVDLKEFGFTPQPSLPSLSRSE--CLNIPKLPVM
Dolphin                 PGNRECKHHCKNKHTC-GHDCCKNGVAQKSEMKESAVSSYLSDLRNRNAVSSLPPVKRLKMQMNKSQSVDLKEFGFTPKPSLTSISRSE--YLNIPELPIM
Killer whale            PGNRECKHHCKNKHTC-GHDCCKNGVAQKSEMKESAVSSYLSDLRNRNAVSSLPPVKRLKMQMNKSQSVDLKEFGFTPKPSLTSISRSE--YLNIPELPIM
Cow                     PGNRECNHHCKNKHTC-GHDCCKTGVAQKSEGKESTISSYLSDLRNRNAVSSLPPVKRLKMQMNKSPSVDLKDFGFAPKPSLPSIARSE--YLNTPELPIM
Tibetan antelope        PGNRECNHHCKNKHTC-GHDCCKTGVAQKSEVKESTISSYLSDLRNRNAVSSLPPVKRLKLQVNKSPGVDLKDFGFAPKPSLPSIARSE--YLNTPELPVM
Sheep                   PGNRECNHHCKNKHTC-GHDCCKTGVAQKSEVKESTISSYLSDLRNRNAVSSLPPVKRLKMQVNKSPGVGLKDFGFAPKPSLPSIARSE--YLNTPELPVM
Domestic goat           PGNRECNHHCKNKHTC-GHDCCKTGVAQKSEVKESTISSYLSDLRNRNAVSSLPPVKRLKMQVNKSPGVDLKDFGFAPKPSLPSIARSE--YLNTPELPVM
Human                   PGNRECNHLCKSKHTC-GHDCCKIGVAQKSEIKESTISSYLSDLRNRNAVSSVPPVKRLKIQMNKSQSVDLKEFGFTPKPSLPSISRSE--YLNISELPIM
Horse                   PGNRECNHHCKNKQTC-GHDCCKTGVAQKSEMKETTISLYLSDLRNRNAVSSLPPVKRLKMQMNKSQSVDLKEFGFTAKPSLPSVSRSE--YLNIPELSVM
White rhinoceros        PGNRECNHHCKNKHTC-GHDCCKTGVAQKSEMKEPTISSYLSDLRNRNAISSLPPVKRLKMQMNKSQNVDLKEFGFSPKPSLPSVSRSE--YLNLPELSIM
Cat                     PGNRECNHHCKNKHTC-GHDCCKTGVAQKSVMKESTISSYLSDLRNRNAVSSLPPVKRLKMQMNKSQSVDLKEFGFTPKPSLSSISRSE--CLNIPELSIM
Dog                     PGNRECNHHCKNKHTC-GHDCCKTGVAQKSETKEPTISSYLSDLRNRNAVSSLPPVKRLKMQMNKSRSVDLKEFGFTPKPSVPSISRSE--CLNIPGLSVM
Ferret                  PGNRECNHHCKNKHTC-GHDCCKTGVAQRSEMKEPTISSYLSDLRNRNAVSSLPPVKRLKMQMNKSQSVDLKEFCFTPKPSLPSISRSE--CLNIPELSIM
Panda                   PGNRECNHHCKNKRTC-GHDCCKTGVAQKSEIKEPTISSYLSDLRSRNAVSSLPPVKRLKMQMNKSQSVDLKEFGFTPKPSLPSISRSE--CLNIPELSIM
Pacific walrus          PGNRECNHHCKNKHTC-GHDCCKTGVAQKSEMKEPTISSYLSDLRNRNAVSSLPPVKRLKMQMSKSQSVDLKEFGFTPKPSLPSISRSE--CLNIPELSIM
Weddell seal            PGNRECNHHCKNKHTC-GHDCCKTGVAQKSEMKEPTISSYLSDLRNRNAVSSLPPVKRLKMQMNKSRSVDLKEFGFTPKPSLPSISRSE--CLNIPELSIM
```

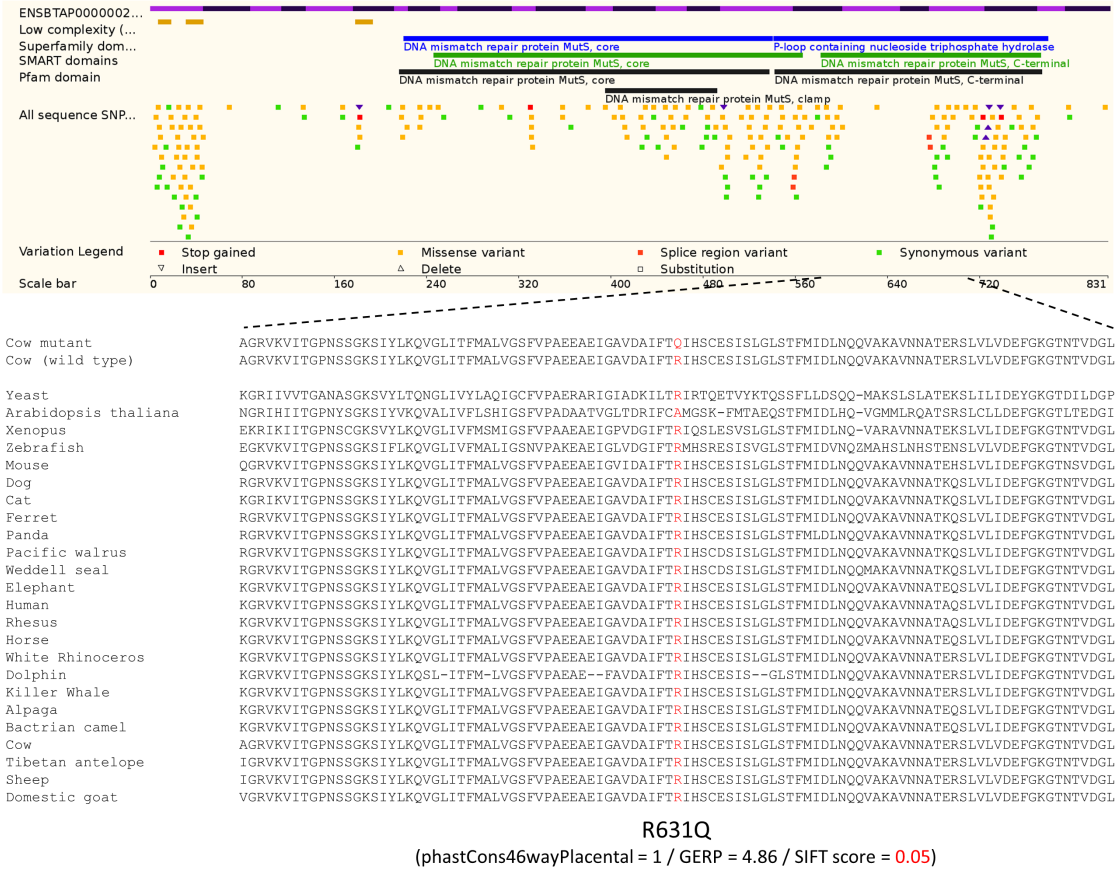
S1189L

(phastCons46wayPlacental = 1 / GERP = 5.41 / SIFT score = 0.62)

Supplemental Figure S22. The S1189L mutation in *HFM1*. Known protein domains are indicated along the bovine protein (source: Ensembl). Alignment of the protein around S1189L missense mutation mainly in placental mammals. SIFT scores and PhastCons and GERP evolutionary scores are also reported.

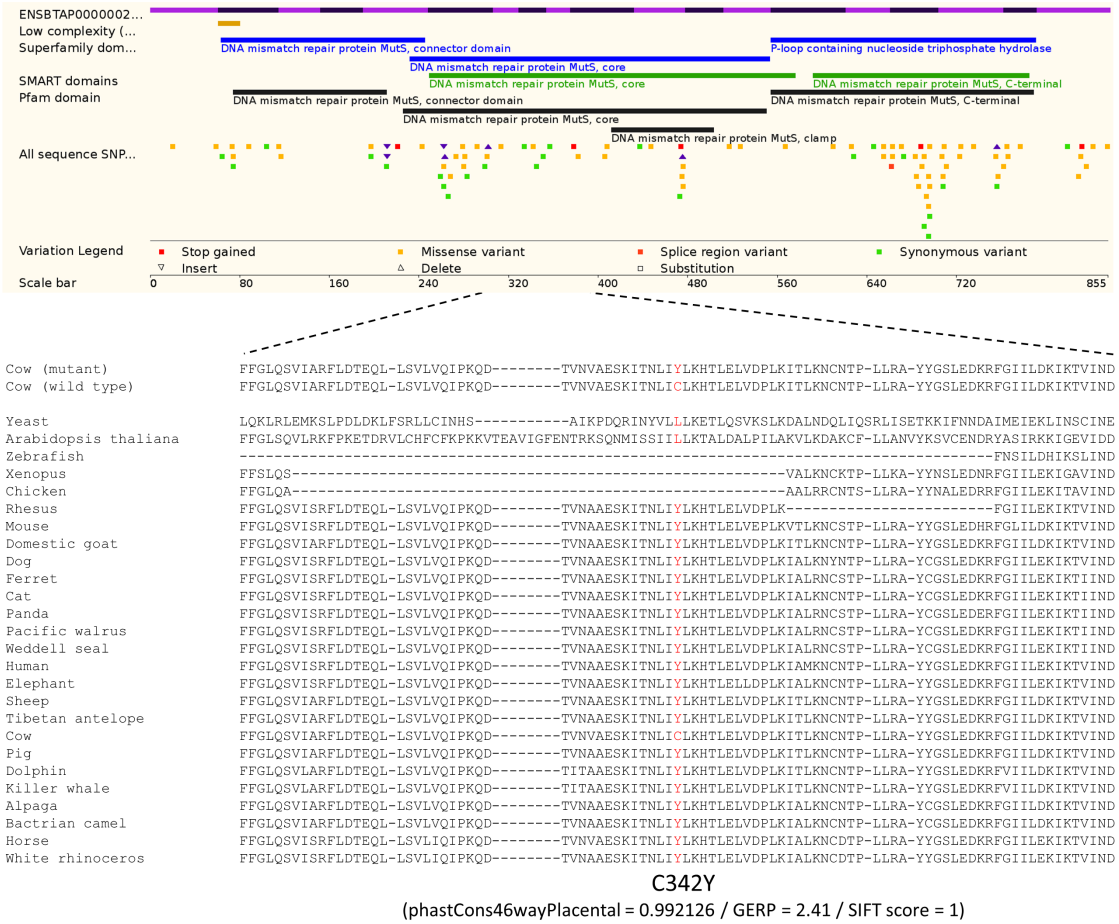


Supplemental Figure S23. The N408S mutation in *MLH3*. Known protein domains are indicated along the bovine protein (source: Ensembl). Alignment of the protein around N408S missense mutation mainly in placental mammals. SIFT scores and PhastCons and GERP evolutionary scores are also reported.

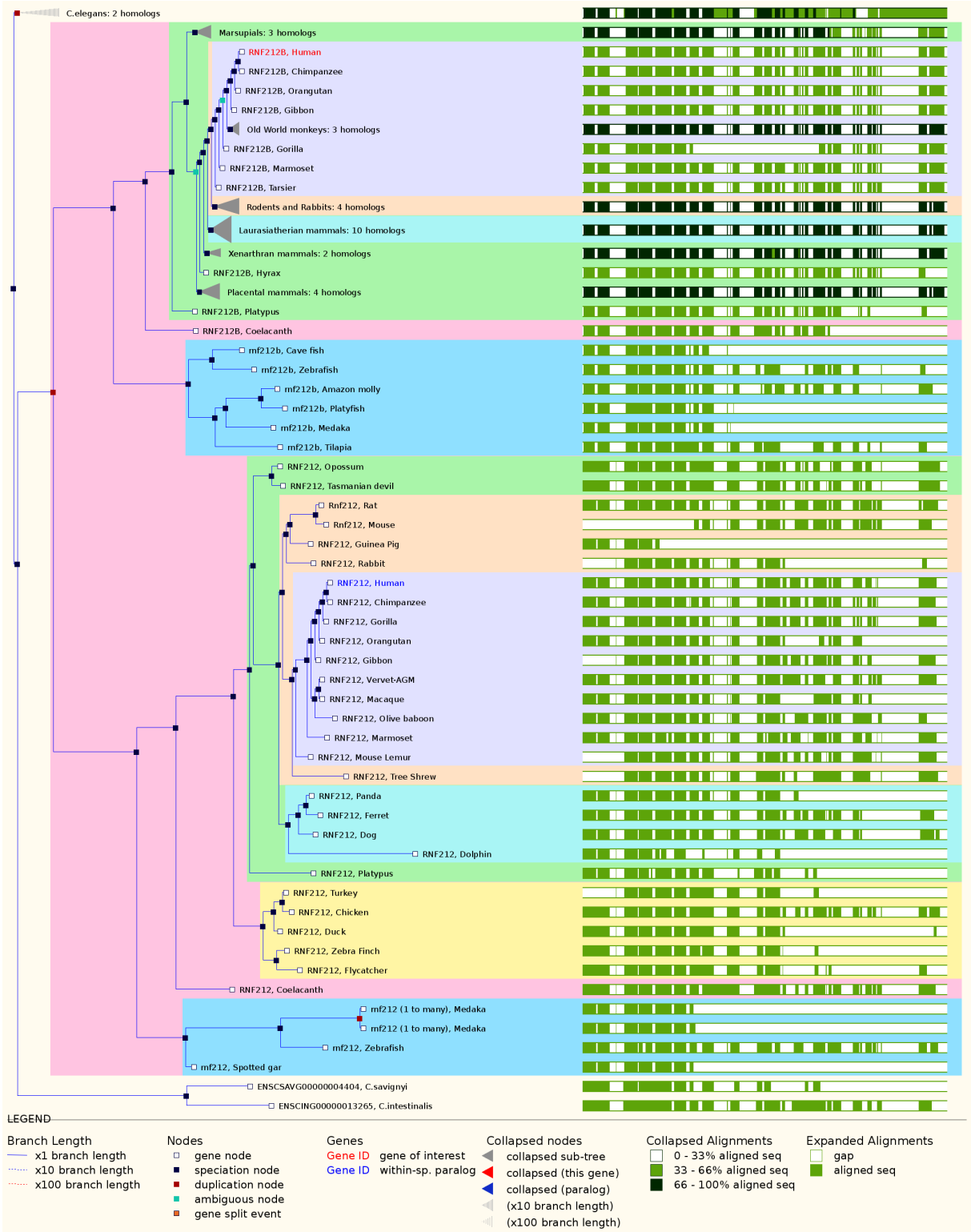


Supplemental Figure S24. The R631Q mutation in *MSH5*. Known protein domains are indicated along the bovine protein (source: Ensembl). Alignment of the protein around R631Q missense mutation mainly in placental mammals. SIFT scores and PhastCons and GERP evolutionary scores are also reported.

Supplemental Figure S25. The A77T mutation in *RNF212*. Known protein domains are indicated along the human protein (source: Ensembl). Alignment of the protein around A77T missense mutation mainly in placental mammals. PhastCons and GERP evolutionary scores are also reported



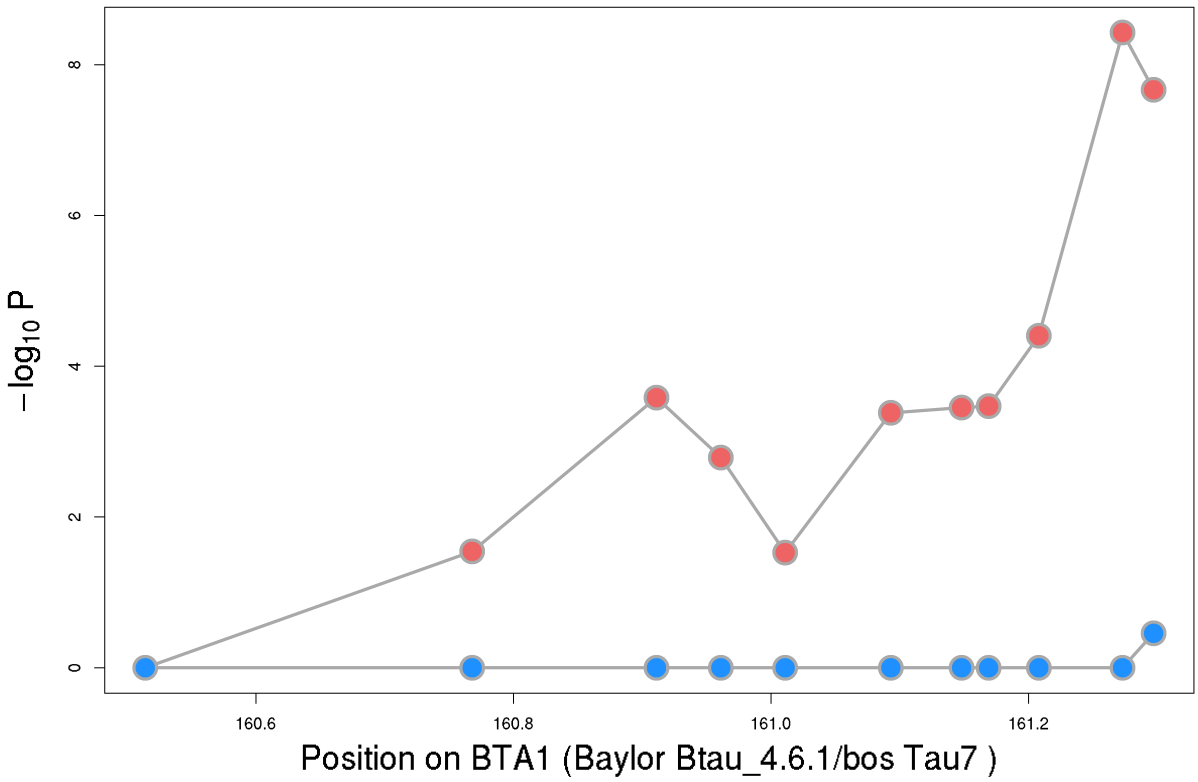
Supplemental Figure S26. The C342Y mutation in *MSH4*. Known protein domains are indicated along the bovine protein (source: Ensembl). Alignment of the protein around C342Y missense mutation mainly in placental mammals. SIFT scores and PhastCons and GERP evolutionary scores are also reported.



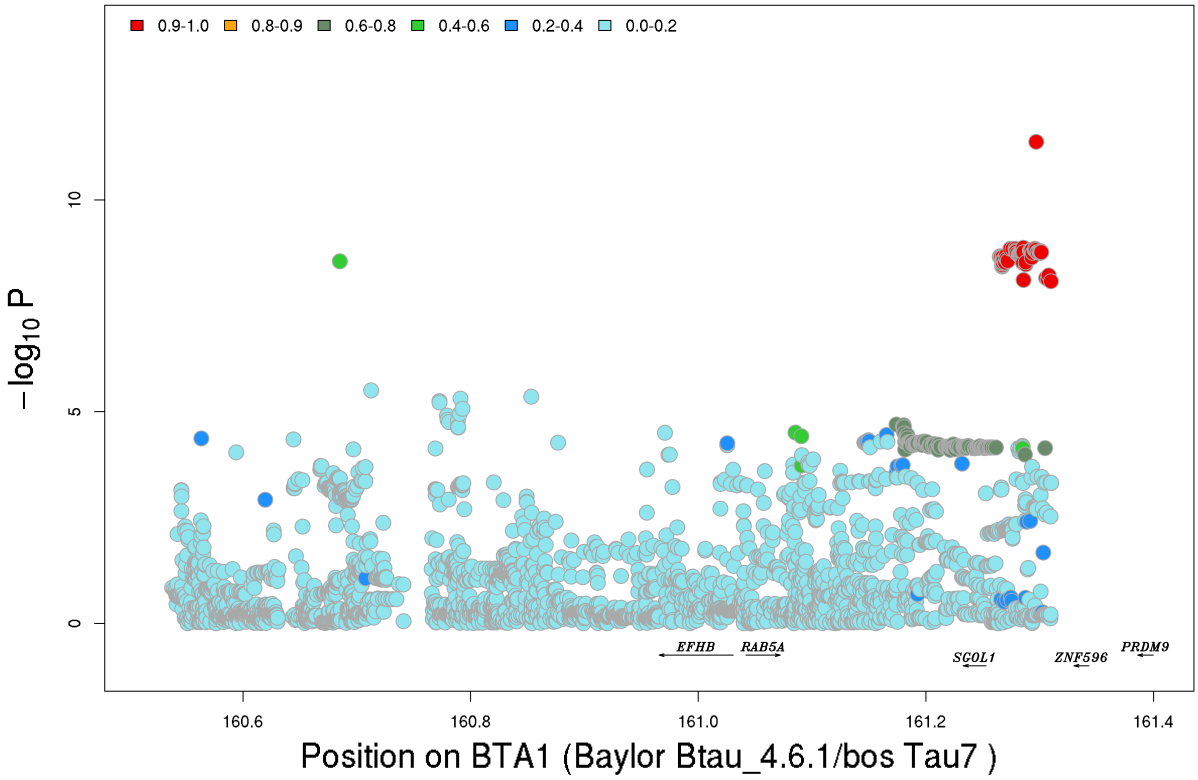
Supplemental Figure S27. Gene family tree of RNF212 and RNF212B (source: Ensembl).

The figure was downloaded from:

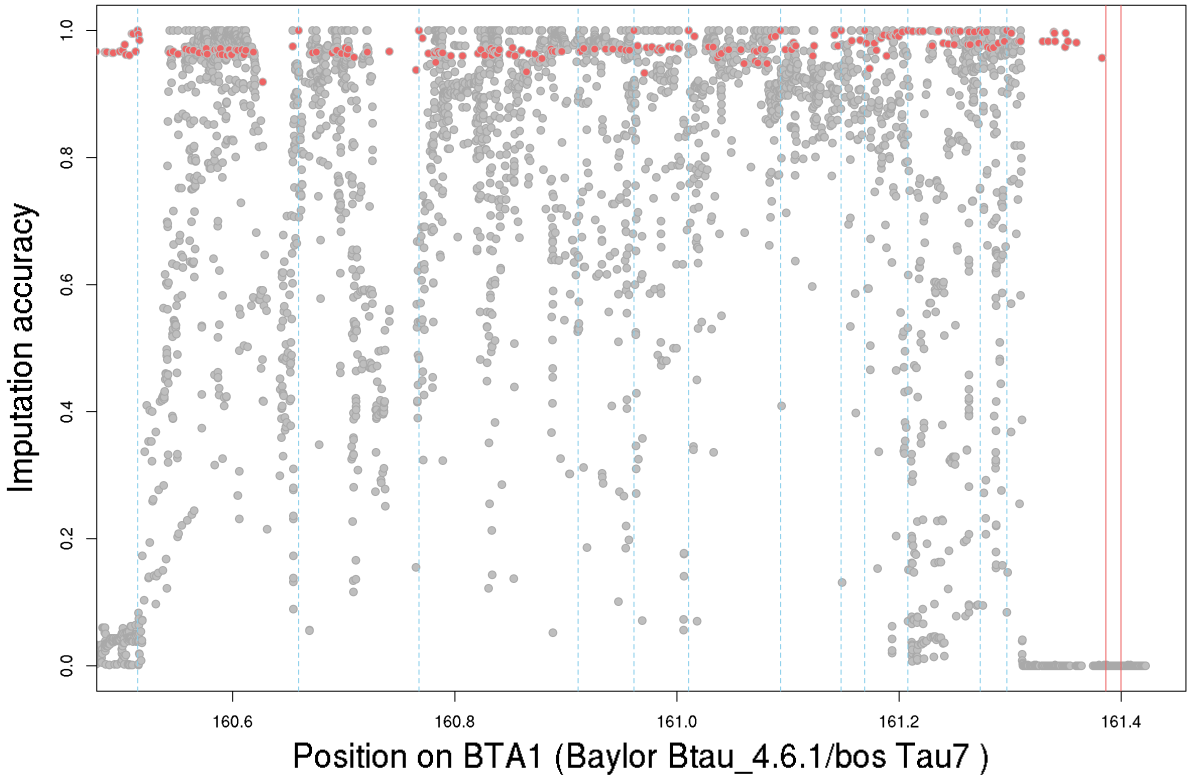
http://www.ensembl.org/Homo_sapiens/Gene/Compara_Tree?db=core;g=ENSG00000215277;r=14:23185316-23273477;collapse=11143267,11143248,11143237,11143242,11143223,11143190,11143185



Supplemental Figure S28. Haplotype-based association with GRR in males (blue) and females (red), for the *PRDM9*-containing BTA1 segments that were misplaced in the UMD3.1 bovine genome build (positions from 44.510 to 45.127 Mb and from 138.433 to 138.828 Mb) and that should map to the telomeric region of BTA1. We relied on the Baylor 4.6.1 bovine genome build (in which these two segments are correctly located) to perform the haplotype based association in the region (160.4 – 161.4 Mb).



Supplemental Figure S29. Sequence-based singe SNP association analysis in females for the *PRDM9*-containing BTA1 segment using the Baylor 4.6.1 bovine genome build (160.5 – 161.4 Mb). Imputation was conducted using Beagle. Variants are colored according to their LD with the lead variant (rs110203897).



Supplemental Figure S30. Imputation accuracy for the *PRDM9*-containing BTA1 segment using the Baylor 4.6.1 bovine genome build (160.5 – 161.4 Mb)**.** Imputation was performed in two steps: from the Illumina 50K array to the Illumina BovineHD array and then to the sequence level. The vertical blue lines mark the position of genotyped SNPs interrogated by the Illumina 50K array. Red dots correspond to imputed SNPs interrogated by the Illumina BovineHD array (imputed in the first step with the larger reference population). Gray dots correspond to imputed SNPs from the sequence data. The vertical red lines mark the limits of the *PRDM9* paralogue.

Supplemental Table S1. List of genes (Homo sapiens) related to reciprocal meiotic recombination - GO:0007131 (source: GO Ontology database released 2015-08-06)

Gene name	Gene symbol
Meiosis-specific with OB domain-containing protein	MEIOB
Double-strand break repair protein MRE11A	MRE11A
DNA repair protein RAD51 homolog 3	RAD51C
Serine-protein kinase ATM	ATM
DNA topoisomerase 2-beta	TOP2B
DNA repair and recombination protein RAD54B	RAD54B
E3 ubiquitin-protein ligase CCNB1IP1	CCNB1IP1
DNA mismatch repair protein Msh6	MSH6
Pachytene checkpoint protein 2 homolog	TRIP13
DNA mismatch repair protein Mlh1	MLH1
Meiotic recombination protein REC8 homolog	REC8
Ubiquitin-conjugating enzyme E2 B	UBE2B
DNA repair endonuclease XPF	ERCC4
Fanconi anemia group M protein	FANCM
Probable ATP-dependent DNA helicase HFM1	HFM1
Testis-expressed sequence 11 protein	TEX11
DNA topoisomerase 2-alpha	TOP2A
Synaptonemal complex central element protein 3	SYCE3
DNA repair protein RAD51 homolog 1	RAD51
DNA repair protein RAD50	RAD50
Probable E3 SUMO-protein ligase RNF212	RNF212
Cyclin N-terminal domain-containing protein 1	CNTD1
Double-strand-break repair protein rad21 homolog	RAD21
Stimulated by retinoic acid gene 8 protein homolog	STRA8
Kelch domain-containing protein 3	KLHDC3
DNA repair protein RAD51 homolog 4	RAD51D
Centromere protein S	APITD1
Synaptonemal complex protein 1	SYCP1
DNA repair protein RAD51 homolog 2	RAD51B
DNA mismatch repair protein Mlh3	MLH3
Meiotic recombination protein SPO11	SPO11
MutS protein homolog 5	MSH5
Centromere protein X	STRA13
MutS protein homolog 4	MSH4
MutS protein homolog 2	MSH2
Meiotic recombination protein DMC1/LIM15 homolog	DMC1
DNA mismatch repair protein Msh3	MSH3

Supplemental Table S2. List of genes spanned by LD-based sets of candidate variants for each QTL obtained using the UMD3.1 genome assembly and the Ensembl gene predictions (version 83).

QTL	Signal	Genes with a variant in the credible set	Additional genes spanned by credible variants (+/ 100 kb)
BTA3 (52.4 Mb)	Primary	*HFM1*	*CDC7L1, LRRC8D, LRRC8C, ENSBTAG00000046077*
BTA6 (122.4Mb)	Primary	*RNF212, FGFRL1*	*GAK, TMEM175, DGKQ, IDUA, SLC26A1, ENSBTAG00000002247, SPON2, CTBP1*
	Secondary	*POLN, NELFA, RNF212*	*WHSC1, C4orf48, SLC2A9, HAUS3, NAT8L, PIGG, DGKQ, IDUA, SLC26A1, FGFRL1, ENSBTAG00000002247, SPON2, CTBP1*
BTA10 (21.2 Mb)	Primary	*PABPN1, BCL2L2, PPP1R3E, KIAA1443 (HOMEZ), RNF212B*	*MYH7, CMTM5, IL25, EFS, SLC22A17, SLC7A8, ENSBTAG00000040053*
	Secondary – male	*RNF212B*	*EFS, SLC22A17, PABPN1, BCL2L2, PPP1R3E, KIAA1443 (HOMEZ), SLC7A8, CEBPE*
	Secondary - female	*RNF212B*	*EFS, SLC22A17, PABPN1, BCL2L2, PPP1R3E, KIAA1443 (HOMEZ), SLC7A8, CEBPE*
BTA10 (86.51Mb)	Primary	*EIF2B2, MLH3, ACYP1, ZC2HC1C, NEK9, TMED10*	*DLST, RPS6KL1, PGF*
BTA18 (52.85Mb)	Primary	*TMEM160, NPAS1*	*ARHGAP35, ZC3H4, SAE1*
BTA23 (27.22Mb)	Primary	*SKIV2L, CFB, C2, ZBTB12, SLC44A4, MSH5, ABHD16A, AIF1, LTB, NFKBIL1, ATP6V1G2*	*TNXB, ENSBTAG00000006864, CYP21A2, C4A, DXO, NELFE, EHMT2 , NEU1, HSPA1L, HSPA1A, LSM2, VWA7, VARS, CLIC1, DDAH2, LY6G6C, C6ORF25, LY6G6D, LY6G6E, LY6G6F, LY6G5B, CSNK2B, GPANK1, C23H6ORF47, BAG6, APOM, PRRC2A, NCR3, TNF, LTA, DDX39B, MGC126945, ENSBTAG00000031905, BOLA-A, MIC1, STK19, SAPCD1, LY6G5C, MCCD1*
BTA3 (69.42)	Primary	*MSH4, ASB17*	*RABGGTB, ACADM*

Supplemental Table S3. PANTHER Overrepresentation Test (release 20150430) for the list of 104 protein coding genes spanned by the LD-based sets of candidate variants (Annotation Version and Release Data: GO Ontology database Released 2015-08-06 / Reference List: Homo sapiens / Annotation Data Set: GO biological process complete).

GO biological process complete	Number of genes in reference list	Number of genes in analyzed list	Expected	Fold Enrichment	+/-	P value
Chiasma assembly	7	3	0.03	>5	+	3.87E-02
Synapsis	38	6	0.17	>5	+	2.04E-04
Reciprocal meiotic recombination	37	5	0.17	>5	+	6.64E-03
Reciprocal DNA recombination	37	5	0.17	>5	+	6.64E-03
Homologous chromosome segregation	45	6	0.20	>5	+	5.49E-04
Chromosome organization involved in meiosis	48	6	0.22	>5	+	8.00E-04
Meiotic chromosome segregation	63	7	0.28	>5	+	1.47E-04
Regulation of humoral immune response	47	5	0.21	>5	+	2.12E-02
Meiosis I	81	7	0.37	>5	+	8.00E-04
Meiotic nuclear division	134	7	0.61	>5	+	2.24E-02
Nuclear chromosome segregation	135	7	0.61	>5	+	2.35E-02
Meiotic cell cycle process	149	7	0.63	>5	+	2.97E-02

Supplemental Table S4. Frequencies of identified variants in genotyped populations.

Variants affecting recombination rate	French Holstein		Dutch Holstein		Holstein-Friesian (New-Zealand)		Jersey (New-Zealand)	
	Males	Females	Males	Females	Males	Females	Males	Females
HFM1 S1189L	0.23	0.21	0.14	0.18	0.15	0.14	0.12	0.12
rs210318688	0.26	0.25	0.22	0.21	0.16	0.15	0.01	0.01
RNF212 P259S	0.26	0.22	0.32	0.24	0.31	0.32	0.08	0.10
RNF212 A77T	0.00	0.00	0.00	0.00	0.08	0.07	0.00	0.00
rs381356614	0.18	0.13	0.22	0.14	0.16	0.14	0.05	0.04
rs207682689	0.64	0.63	0.56	0.61	0.58	0.57	0.30	0.30
rs437013002	0.00	0.00	0.00	0.00	0.04	0.06	0.18	0.16
MLH3 N408S	0.50	0.50	0.42	0.44	0.57	0.56	0.03	0.04
MSH5 R631Q	0.06	0.08	0.13	0.07	0.10	0.08	0.00	0.00
rs135941180	0.53	0.59	0.60	0.51	0.54	0.58	0.66	0.65

Supplemental Table S5. Number of parents (males and females) genotyped in different populations and their respective number of genotyped progeny (records).

Populations	Males		Females	
	Parents	Records	Parents	Records
All	2940	94516	11461	25332
Dutch Holstein	233	5080	322	970
French Holstein	664	40821	5451	16912
Holstein-Friesian	891	19398	907	1399
Jersey	589	19028	1301	1745
Population from New-Zealand*	2043	48615	5688	7450

*Includes the crossbred individuals

Supplemental Note S1. Distribution of global recombination rate per sex.

The distribution of global recombination rate (GRR) in males and females is plotted in Figure 1 for two populations. In Table 1, mean GRR per sex and male to female ratio are reported for all the populations used in the present study.

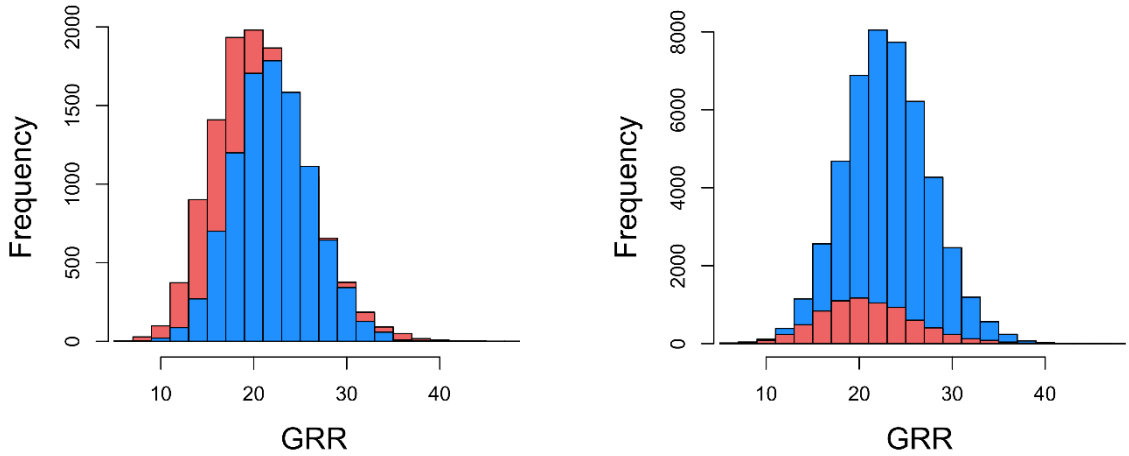


Figure 1. Distribution of number of CO per gamete (or offspring) from male (blue) or female (red) parents. The left panel compares distributions in the French Holstein population (9,651 males; 12,557 females) and the right panel was obtained with data from the population from New-Zealand (46,664 males, 7,431 females).

Table 1. Global recombination rate in males (M) and females (F) measured as the number of crossovers per gamete (Mean and Variance). Also provided is the M/F ratio and corresponding standard error (SE)

Population	Sex	Number of Records	Mean	Variance	Ratio (M/F)	SE (Ratio)
	M	59427	23.3	20.9		
All populations	F	20729	21.4	25.2	1.09	0.002
	M	9651	22.6	17.6		
French Holstein	F	12557	21.4	24.4	1.06	0.003
	M	3112	23.3	20.9		
Dutch Holstein	F	741	21.6	22.9	1.08	0.010
	M	46664	23.5	21.3		
New-Zealand population	F	7431	21.5	26.8	1.09	0.003
	M	18699	24.7	21.8		
Holstein-Friesian (New-Zealand)	F	1408	22.6	29.1	1.10	0.007
	M	17896	22.0	17.8		
Jersey (New-Zealand)	F	1725	20.0	22.6	1.10	0.006

Sex comparison of number of CO per chromosome reveals that the higher recombination rate in males in consistent across chromosomes (Figure 2).

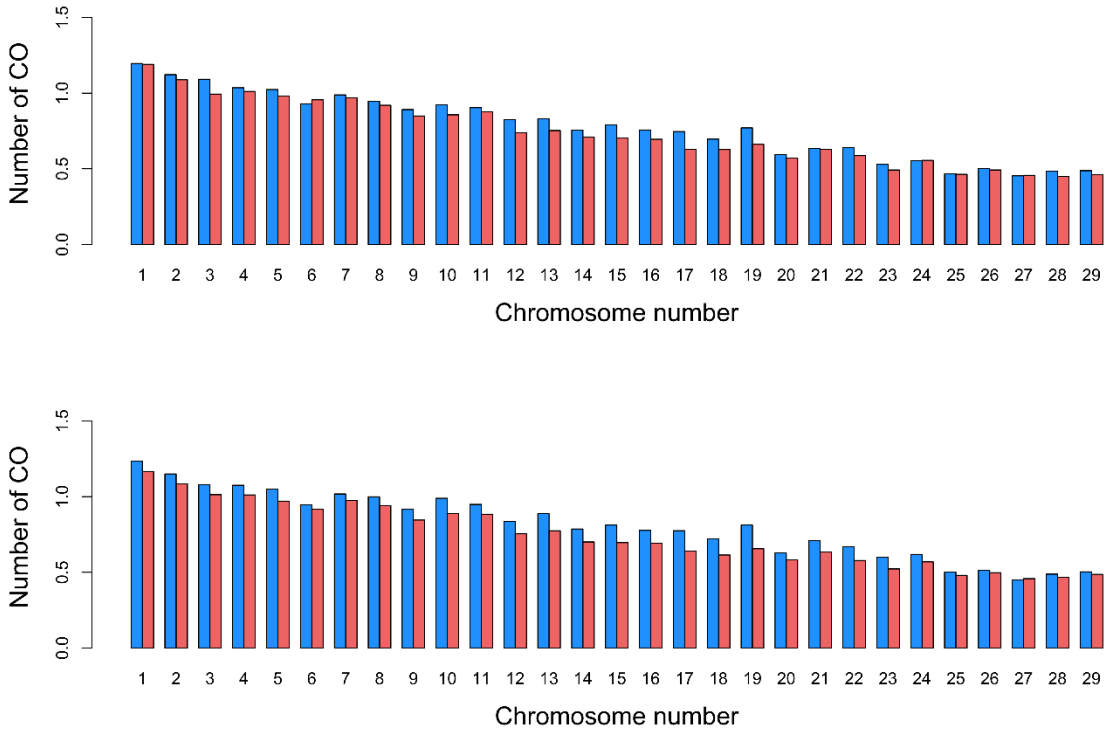


Figure 2. Number of CO in males (blue) and female (red) observed for the 29 bovine autosomes in the French Holstein population (top) and the population from New-Zealand (lower panel).

Supplemental Note S2. Imputation and mapping of the RNF212 *A77T* variant

When fitting the P259S variant in the haplotype-based association model (Supplementary Figure 8), a highly significant QTL was still present in males (p = 3.6e-14) whereas the significance was lower in females (p = 1.3e-2 in the QTL region). This signal is driven by one Ancestral Haplotype (hereafter called AHAP14) that has the most negative effect on GRR although carrying the P259S NS mutation (Fig. 1). This suggests that a second variant with a negative effect is associated to that haplotype (D' = 1 between the two variants).

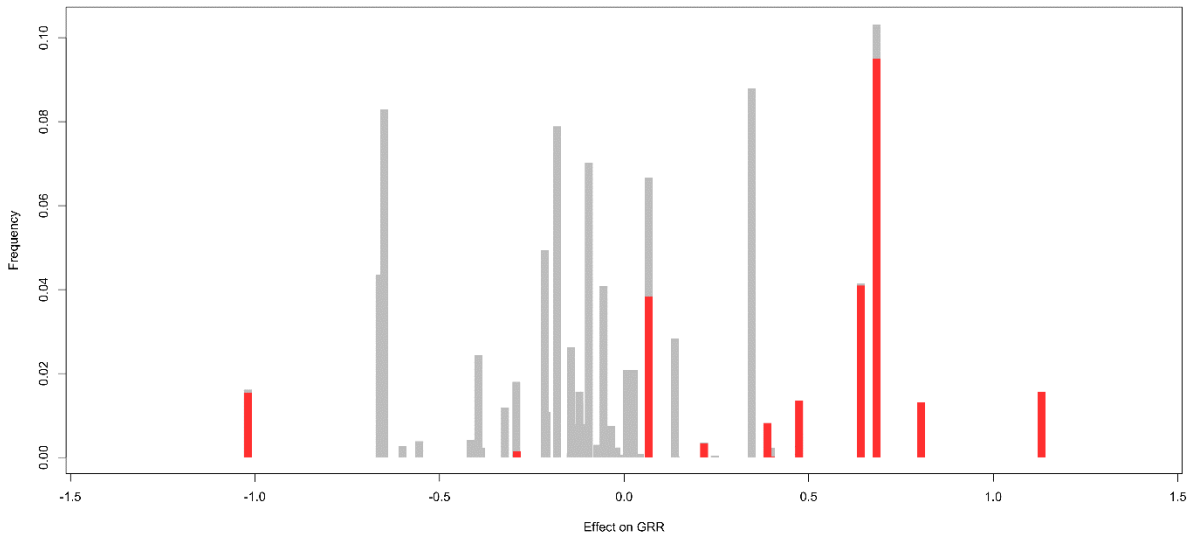


Figure 1. Frequency and effects on male GRR of the 50 AHAP (Ancestral Haplotypes). The estimated frequency of the *RNF212* P259S variant per AHAP is represented in red. AHAP perfectly associated with the P259S mutation present the strongest effect except for one AHAP (the AHAP14) with a very negative effect on GRR (-1 GRR).

AHAP14 segregates only in Holstein-Friesian from New-Zealand (~6%), therefore we repeated the sequence-based association analysis in males from New-Zealand including the P259S variant as a covariate (and variants identified for the primary signals). The association (Figure 2, Suppl. File 1) maximized in a cluster of 20 SNPs located from chr6:118983781 to chr6:119149263 (p-val = 3.8e-16 at position chr6:119117260), 785 kb from *RNF212* (JZ078754), and encompassing *C4orf48, NAT8L* and *Poln*, not

previously associated to meiosis (although *Poln* is a DNA polymerase involved in DNA repair and homologous recombination). All the nine sequenced carriers of AHAP14 (out of 91 individuals from New-Zealand that are both sequenced and genotyped) carry one copy for each of these variants (seventeen of them being in perfect (r^2 = 1) with AHAP14). Among the other variants in high LD with AHAP14 in those individuals, we observe an A77T mutation in *RNF212* (chr6: 118193274). Indeed, ten sires carry that mutation, including the nine carriers of AHAP14 and one carrier of a very similar haplotype (IBS for 18 markers, or a 1.6 Mb segment). In the population of New-Zealand, that variant has a p-value of 2.0e-9 but also a low imputation accuracy (r^2 = 0.60). The variant has a relatively high correlation with the lead variant in the sequence data (r = 0.96 - genotypes are identical for 121 out of 122 sequenced sires), but this association dramatically drops after imputation (r = 0.41). In cattle, lower imputation accuracy has been reported for rare variants (e.g., Druet et al. 2014; van Binsbergen et al. 2014). In addition, imputation accuracy is affected by border effects (we are close to the telomere) and by build errors known to be present in that region of the bovine genome assembly. Imputation could therefore be suboptimal in the region of the *RNF212* region and we decided to re-impute the P259S and A77T NS variant identified in *RNF212* with a new strategy described below.

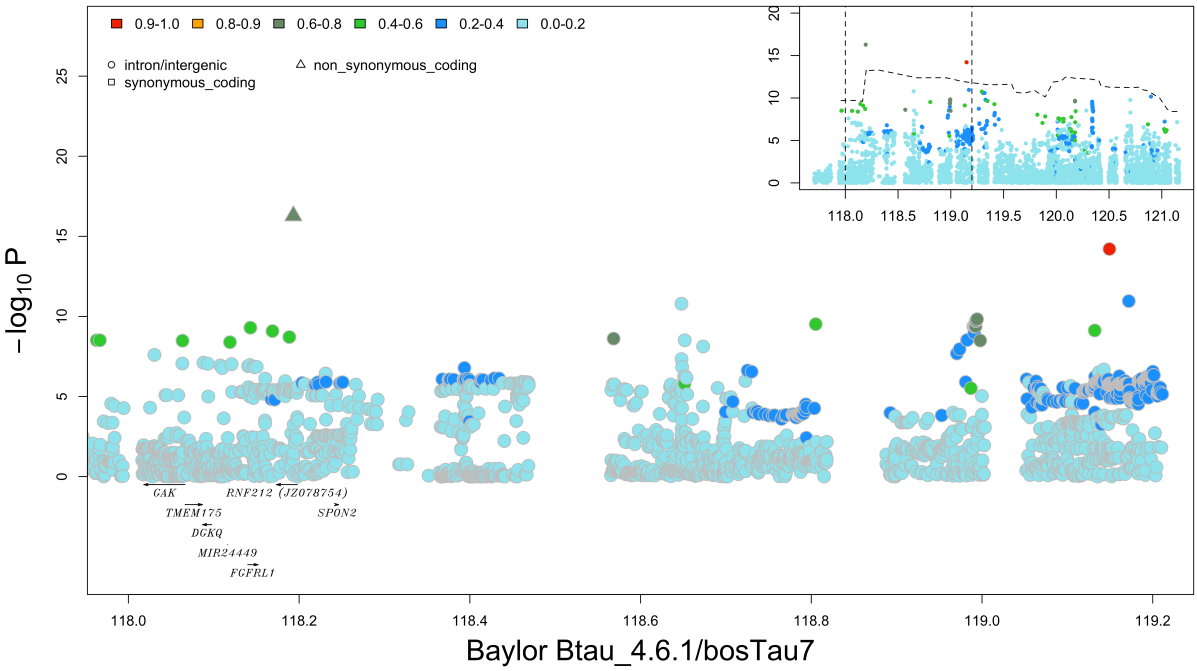


Figure 2. Secondary sequence-based association analysis for the BTA6 QTL associated with *RNF212* with a model including the *RNF212* P259S, the *MLH3* N408S, the *HFM1* S1189L, the *MSH5* R631Q variants and the intronic insertion in *RNF212B* (rs381356614) as covariates. The variant colors represent the LD with the lead variant (the variant with the lowest p-value). The association was performed in males from the population from New-Zealand.

The new strategy relies on the haplotypes obtained with LINKPHASE3 / DAGPHASE and the genotyping arrays. Indeed, these haplotypes are accurate because they are obtained in a much larger population and pedigree information was used to infer them. Many individuals have their parents genotyped (which improves phasing) and LINKPHASE3 is very efficient for animals with many genotyped offspring (which is the case with the genotyping array data). We used the 42 markers in a 3.19 Mb segment (the segment containing *RNF212* and incorrectly mapped in bovine genome assemblies bosTau6 and bosTau7 - see Supplementary Note 4). For the 42 markers, haplotypes obtained from DAGPHASE were unchanged. We then phased each NS variant in the sequenced individuals based on the ancestral haplotypes (AHAP): homozygous individuals were phased de facto whereas for heterozygous individuals we assigned the NS variants to the haplotype associated with an AHAP carrying the variant (there was no conflict). Then we used both phased data set as input for MINIMAC (Howie et al. 2012).

The significance of the A77T mutation increased (p-value 5.3e-17) and was even more significant than the peak of the sequenced based association analysis. Consequently, we added this variant to the credible set.

Supplemental Note S3. Estimation of additive, dominance and epistatic effects

Additive effects. The additive effects of the identified variants (reported in Table 2) were estimated with a linear mixed model (**LMM**) similar to the one used for association study:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Pc} + \mathbf{Z_u u} + \mathbf{Z_p p} + \sum_{i=1}^{N_v} \mathbf{v}_i a_i + \mathbf{e}$$

where **y** is the vector of GRR records, **1** is a vector one ones, μ is the overall mean, **P** and $\mathbf{Z}_u$ and $\mathbf{Z}_p$ are incidence matrices relating phenotypes to principal components (**PC**), polygenic effects and permanent environment effects **c** is a vector with the effects of the first four PC, **u** is a vector with individual polygenic effects, **p** is a vector with permanent environment effects, N_v is the number of variants, $\mathbf{v}_i$ is the vector of allelic dosages for variant *i* for the corresponding records, a_i is the allelic effect for the variant *i*, and **e** is a vector of random residual error terms.

Dominance. Dominance was tested for the same variants (excluding the rare *RNF212* A77T variant - MAF=1.7%) by adding a regression on heterozygote probability to the previous LMM:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Pc} + \mathbf{Z_u u} + \sum_{i=1}^{N_v} \mathbf{v}_i a_i + \mathbf{w}_i d_i + \mathbf{e}$$

where $\mathbf{w}_i$ is a vector with probabilities that probands associated to each record are heterozygote for variant *i* and d_i is the associated effect (dominance deviation). The significance of the dominance effect was assessed by a Z-test.

Table 1. Estimated additive and dominance effects for the identified variants affecting global recombination rate (and associated p-values).

HFM1 S1189L	M	-0.78	0.24	-0.55	-1.57	0.1233
	F	-0.36	-0.22	-0.58	-0.72	0.0726
rs210318688	M	-	-	-	-	-
	F	-0.63	0.19	-0.44	-1.26	0.1472
RNF212 P259S	M	0.90	0.23	1.13	1.79	0.0217
	F	0.60	0.05	0.64	1.19	0.6789
rs381356614	M	1.51	-0.39	1.11	3.02	0.0091
	F	0.94	0.00	0.94	1.88	0.9949
rs207682689	M	-0.56	-0.08	-0.64	-1.12	0.2569
	F	-0.12	0.12	0.00	-0.25	0.1024
rs437013002	M	-0.38	-0.04	-0.42	-0.76	0.8329
	F	-1.57	0.43	-1.14	-3.14	0.2030
MLH3 N408S	M	0.73	0.08	0.81	1.45	0.2910
	F	0.45	0.09	0.54	0.91	0.2423
MSH5 R631Q	M	-1.51	0.49	-1.02	-3.03	0.1890
	F	-	-	-	-	-

[1]Sex: M for male and F for female; [2]Dominance effect: measured as deviation from the dosage effect in heterozygotes; [3]Estimated effect of the genotypic classes (obtained from the dose effect and the dominance effect): the 00 effect was set to 0; [4]p-value of the dominance effect.

Only two out of the 8 tested variants were significant for dominance. These dominance effects were not significant in the second gender.

Interactions. To study the interaction between locus A (with possible genotypes A_1A_1, A_1A_2, A_2A_2) and B (with possible genotypes B_1B_1, B_1B_2, B_2B_2), an effect was associated to the nine possible genotypic interaction classes A_iA_j x B_kB_l where *i*, *j*, *k* and *l* indicate the marker allele. The significance was tested by adding this interaction to the previous LMM:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Pc} + \mathbf{Z}_{\mathbf{u}}\mathbf{u} + \sum_{i=1}^{N_v} \mathbf{v}_i a_i + \mathbf{Z}_{\text{gxg}}\mathbf{gxg} + \mathbf{e}$$

where $\mathbf{Z}_{\text{gxg}}$ is an incidence matrix relating records to genotype interaction effect containing for each record the probabilities of the proband to belong to each of the nine genotypic interaction classes and **gxg** are the random effects associated to each of the nine genotypic interaction classes. The significance of the genotypic interaction was tested with a Likelihood Ratio Test (distributed as a chi-square distribution with 1 df) comparing the likelihood of the data assuming a model with versus without interactions for the two tested variants. All genotypic combinations of identified variants were studied, except those with the *RNF212* A77T variant (MAF=1.7%).

Table 2. Significance of interactions effects for tested genotypic combinations

First Variant	Second Variant	p-value (male)	p-value (female)
HFM1 S1189L	*MLH3 N408S*	*0.0021*	1
RNF212 P259S	*MLH3 N408S*	*0.009*	1
HFM1 S1189L	*RNF212 P259S*	*0.032*	1
RNF212 P259S	*MSH5 R631Q*	*0.0421*	-
RNF212 P259S	rs381356614	0.0828	1
RNF212 P259S	rs210318688	-	1
rs381356614	*MSH5 R631Q*	0.1035	-
rs381356614	rs210318688	-	1
MSH5 R631Q	rs207682689	1	-
HFM1 S1189L	rs381356614	0.1594	0.5839
HFM1 S1189L	rs381356614	-	1
HFM1 S1189L	*MSH5 R631Q*	0.2617	-
rs381356614	*MLH3 N408S*	0.2815	1
MLH3 N408S	*MSH5 R631Q*	0.8625	-
HFM1 S1189L	rs210318688	-	0.6892
HFM1 S1189L	rs207682689	0.3994	0.522
MLH3 N408S	rs207682689	1	0.5169
MLH3 N408S	rs437013002	-	0.8795
rs207682689	rs210318688	-	0.7083
RNF212 P259S	rs437013002	-	0.7459
rs381356614	rs437013002	-	0.78
rs210318688	rs437013002	-	0.8875
rs381356614	rs207682689	-	0.9203
MLH3 N408S	rs210318688	-	1
RNF212 P259S	rs207682689	*0.0002*	-

Five out of 33 interactions were significant. For these five significant interactions, the effect of each of the 3 genotypes (A_1A_1, A_1A_2, A_2A_2) at locus A was estimated in the background of 3 genotypes (B_1B_1, B_1B_2, B_2B_2) at locus B, and vice versa (see Figures 1-4). None of these interactions changes the main additive effect of locus A and B.

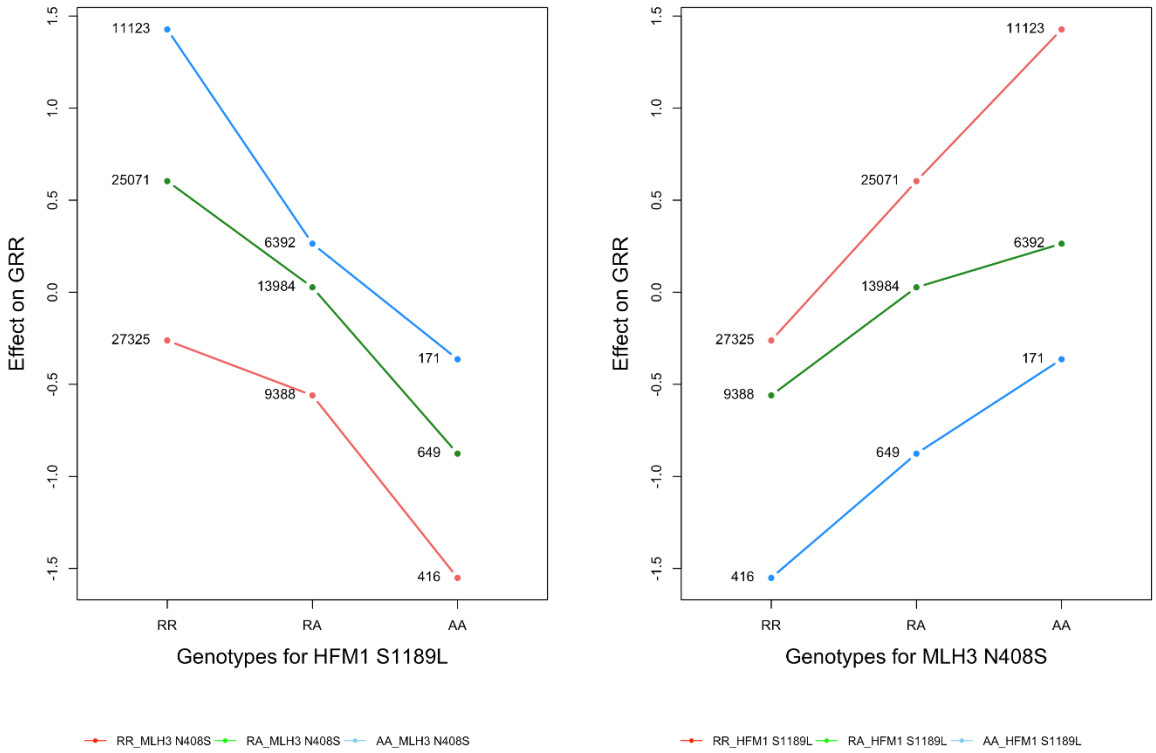


Figure 1. **Interaction between a pair of loci associated to GRR**. The left panel shows the effect of genotypes; Reference Reference [RR], Reference Alternative [RA] and Alternative Alternative [AA] of the *HFM1* S1189L variant on GRR, when the genotypes at the *MLH3* N408S variant are RR (red), RA (green) or AA (blue). Similarly the right panel shows the effect of the *MLH3* N408S variant on GRR in the three genotypic backgrounds of the *HFM1* S1189L variant.

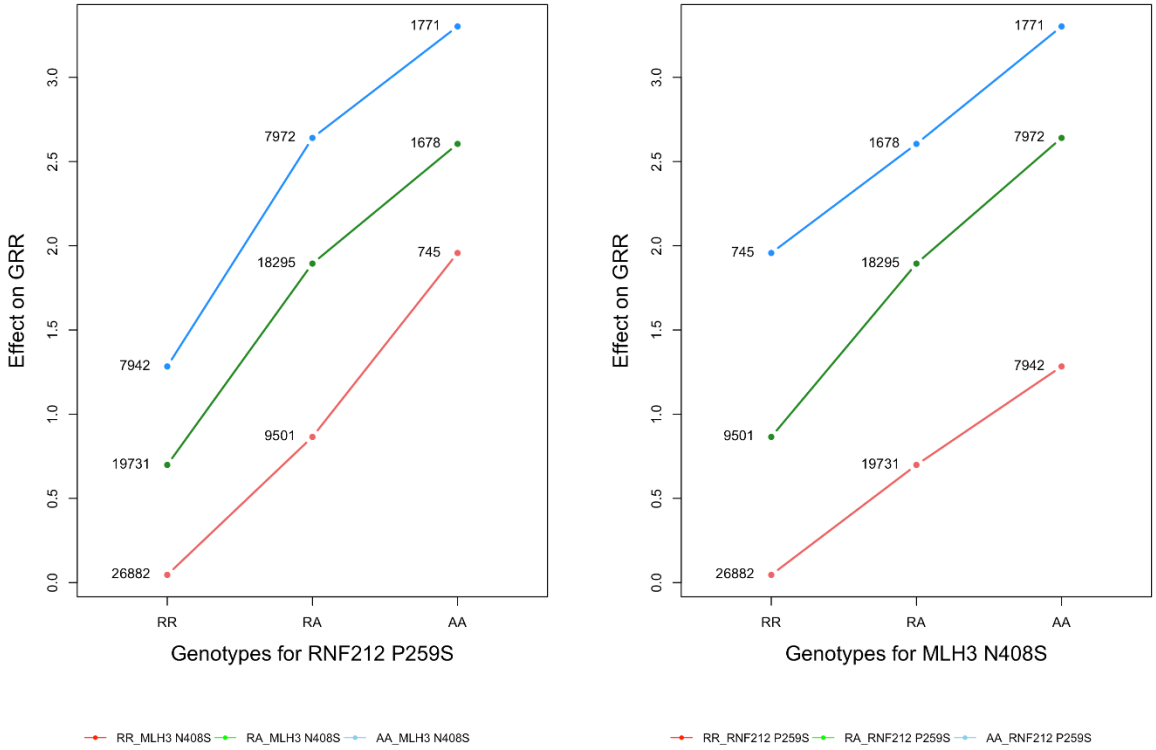


Fig 2. **Interaction between a pair of loci associated to GRR**. The left panel shows the effect of genotypes; Reference Reference [RR], Reference Alternative [RA] and Alternative Alternative [AA] of the *RNF212* P259S variant on GRR, when the genotypes at the *MLH3* N408S variant are RR (red), RA (green) or AA (blue). Similarly the right panel shows the effect of the *MLH3* N408S variant on GRR in the 3 genotypic backgrounds of the *RNF212* P259S variant.

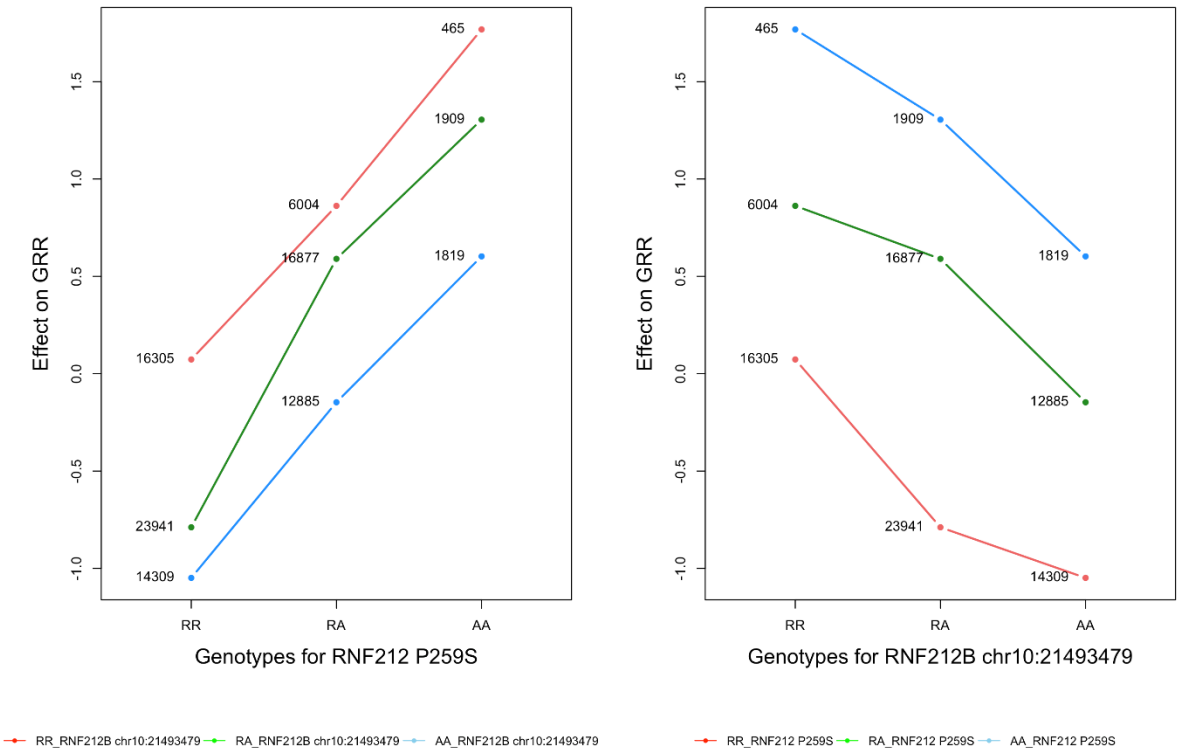


Fig 3. **Interaction between a pair of loci associated to GRR**. The left panel shows the effect of genotypes; Reference Reference [RR], Reference Alternative [RA] and Alternative Alternative [AA] of the *RNF212* P259S variant on GRR, when the genotypes of the rs207682689 chr10:21493479 A to T variant (in *RNF212B*) are RR (red), RA (green) or AA (blue). Similarly the right panel shows the effect of rs207682689 the chr10:21493479 variant (in *RNF212B*) on GRR in the 3 genotypic backgrounds of the *RNF212* P259S variant.

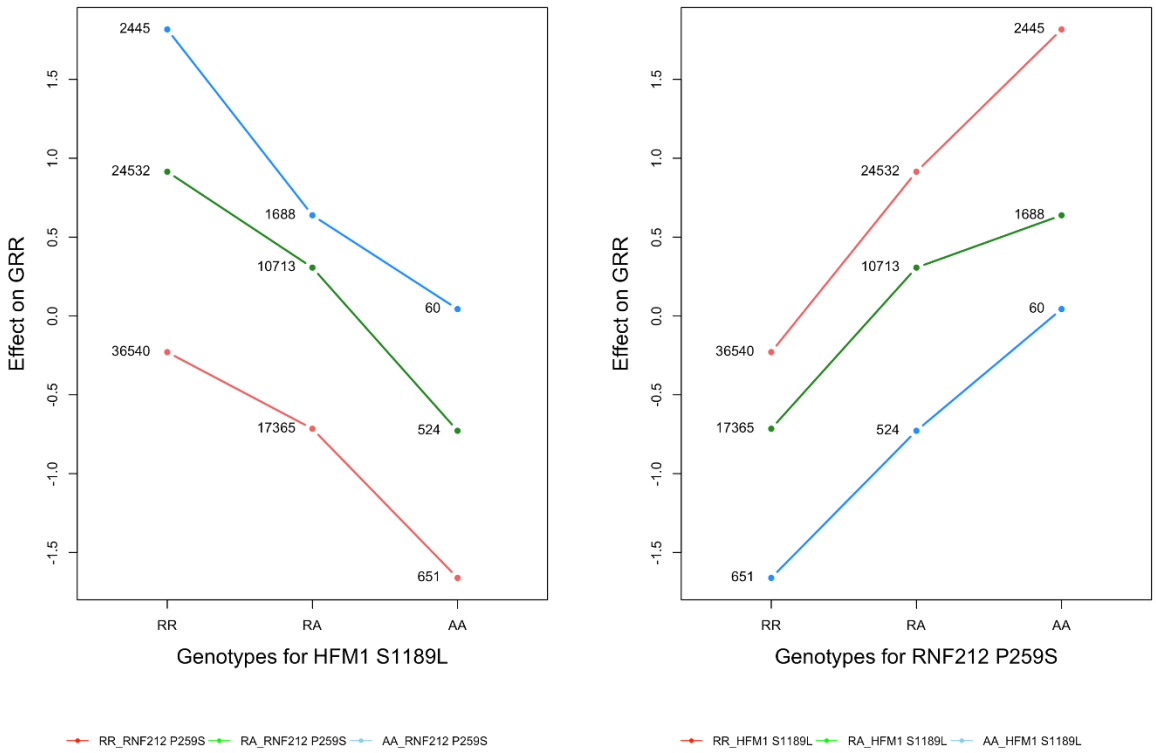


Fig 4 **Interaction between a pair of loci associated to GRR**. The left panel shows the effect of genotypes; Reference Reference [RR], Reference Alternative [RA] and Alternative Alternative [AA] of the *HFM1* S1189L variant on GRR, when the genotypes at the *RNF212* P259S variant are RR (red), RA (green) and AA (blue). Similarly the right panel shows the effect of the *RNF212* P259S variant on GRR in the 3 genotypic backgrounds of *HFM1* S1189L variant.

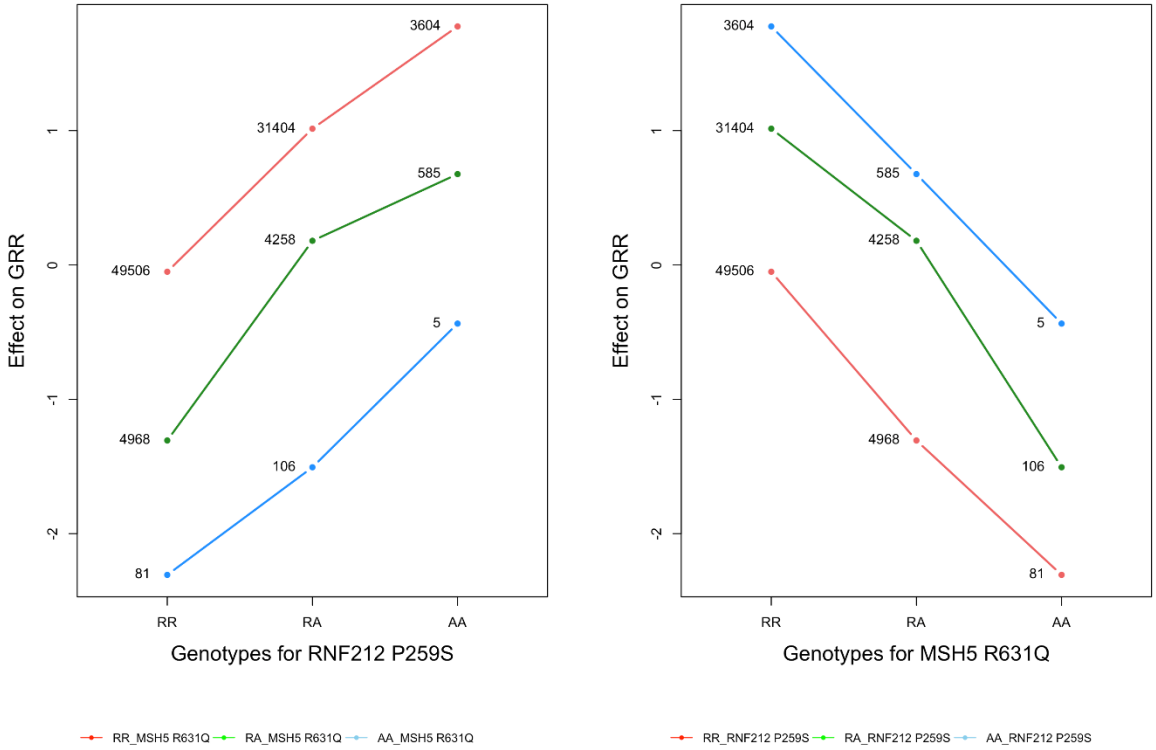


Fig 5. **Interaction between a pair of loci associated to GRR**. The left panel shows the effect of genotypes; Reference Reference [RR], Reference Alternative [RA] and Alternative Alternative [AA] of the *RNF212 P259S* variant on GRR, when the genotypes at the *MSH5 R631Q* variant are RR (red), RA (green) and AA (blue). Similarly the right panel shows the effect of the *RNF212* P259S variant on GRR in the 3 genotypic backgrounds of *MSH5 R631Q* variant.

Supplemental Note S4. Description of marker map correction on BTA6

We have previously shown that LINKPHASE3 can be used to detect putative map errors (Druet and Georges 2015). If a segment is incorrectly mapped, local inheritance patterns (haplotype transmitted by the parent to the progeny) will differ inside the segment compared to the flanking regions. If the segment contains several markers, LINKPHASE3 will detect inflated recombination rate in the flanking marker intervals (excess of double crossing over (**CO**) supported by several markers). If the segment contains one or two markers, the model will rather assume that there is a genotyping error (in either the parent or the progeny) rather than a double CO. Inflated parental genotyping error rate or within haplotype allelic entropy (measuring genotype discrepancies in offspring inheriting the same homolog at a given marker position) will be reported. We defined the Map Confidence Score (**MCS**) as a combination of these three statistics (see Druet and Georges, 2015 for more details).

Using LINKPHASE3, we observed that a large segment on BTA6, ranging from positions 106517120 to 109835442 on Bos_Taurus_UMD3.1 (bosTau6) assembly, was poorly built and incorrectly mapped (see Fig. 1). We also determined by association that the segment should be located at the end of BTA6 (119146065).

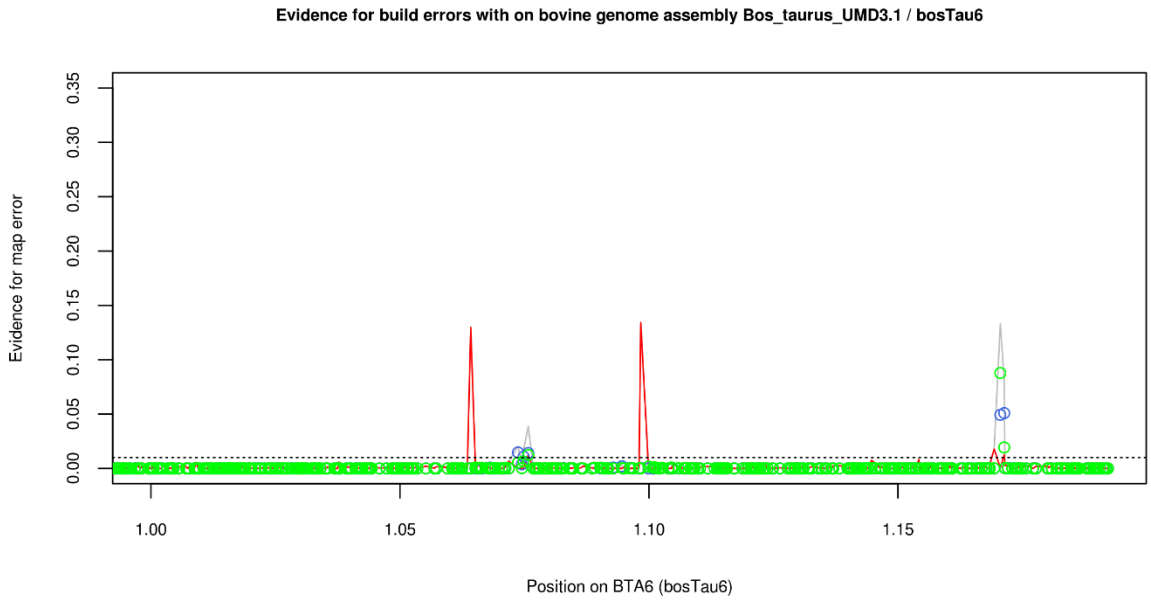


Figure 1. Evidence for errors on BTA6 (bovine assembly bosTau6) after run of LINKPHASE3: estimated recombination rates in marker intervals (red), parent genotyping error rate (blue), within haplotype allelic entropy (green), overall Map Confidence Score (grey).

We used then an alternative bovine genome assembly (Baylor Btau4.6.1 / BosTau7) and ran again LINKPHASE3. We detected less map problems (see Fig. 2) but a segment ranging from positions 117960634 to 121147495 seemed still incorrectly mapped as we observed inflated recombination rates on both sides of the segment.

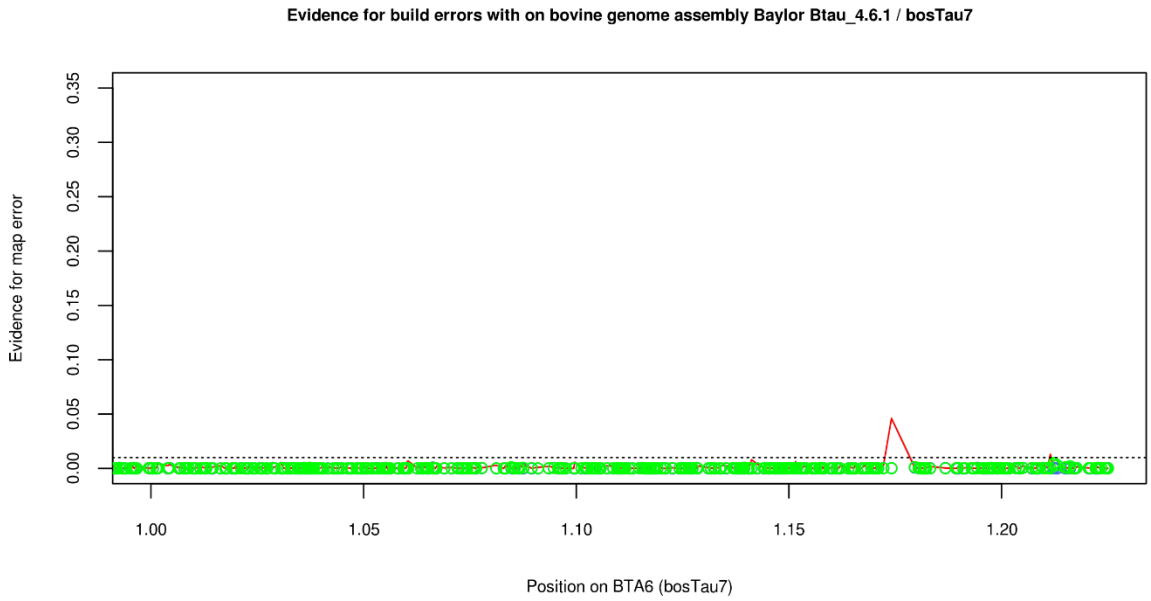


Figure 2. Evidence for errors on BTA6 (bovine assembly bosTau7) after run of LINKPHASE3: estimated recombination rates in marker intervals (red), parent genotyping error rate (blue), within haplotype allelic entropy (green), overall Map Confidence Score (grey).

After performing a liftover from bosTau6 to bosTau7 for the markers on the Illumina Bovine50K genotyping array (Fig 3.), we determined that the same segment was involved in both cases. Local rearrangements within the segment are also visible.

Our first analysis on Bos_Taurus_UMD3.1 assembly indicated us that the segment was incorrectly built, incorrectly mapped and associated to the end of BTA6. With the Baylor Btau4.6.1 assembly, there is no indication of map errors within the segment but its mapping remains erroneous. Based on these observations, we cut the BTA6 marker map obtained with the Baylor Btau_4.6.1 assembly into three segments (positions prior to the segment chr6:0-117411733, the segment chr6:117960634-121147495 and positions posterior to the segment chr6:121218019-122501364) and phased them independently with LINKPHASE3.

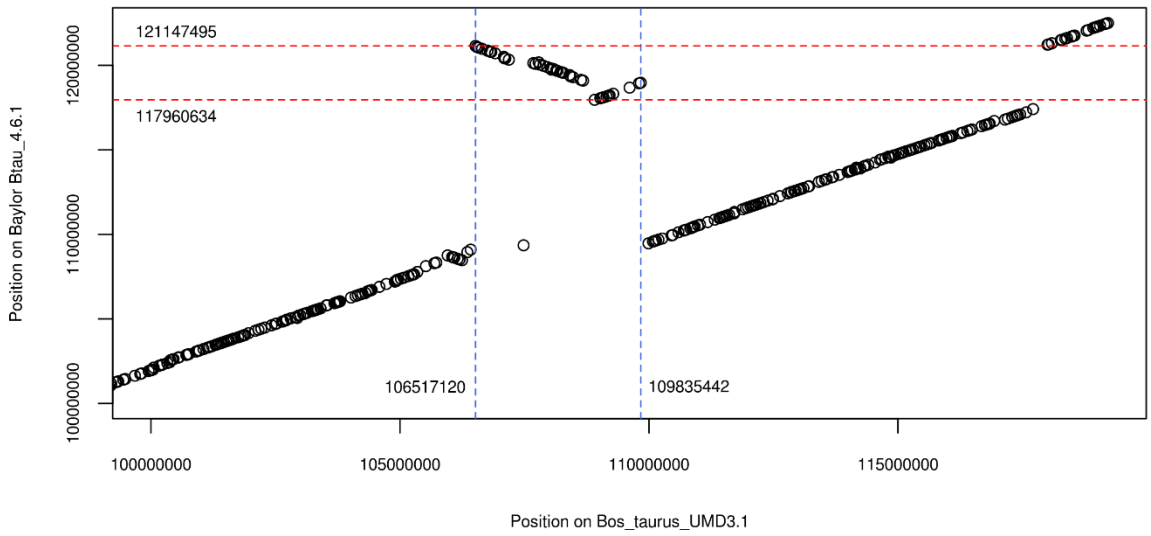


Figure 3. LiftOver between bovine genome assemblies Bos_taurus_UMD3.1 and Baylor Btay4.6.1 for the positions of the markers from the Illumina Bovine50K genotyping array. Dashed lines indicate the limits of the incorrectly mapped segment.

We then compared within families segregation patterns. LINKPHASE3 computes the probability that an offspring inherits the paternal allele of its parent (ranging from 0.00 to 1.00 / 0.00 indicating that the offspring inherited the maternal allele). We kept only meiosis with probabilities smaller than 0.05 or higher than 0.95 (to keep only informative transmissions) and adjusted these inheritance vectors for the mean (0.5). We finally computed the correlation (r) between inheritance vectors. High correlations indicate few CO between tested positions and proximity. Correlations between segments borders are reported in Table 1 and indicate that the chr6:117960634-121147495 segment should be inverted and placed at the end of the chromosome. Indeed, the highest association with 117411733 (end of first segment) is 121218019 (start of third segment) whereas the highest association with 122501364 (end of third segment) is 121147495 (end of second segment).

Table 1. Association between inheritance vectors (measured as r^2) tested at extremities of different segments.

	117960634	121147495	121218019	122501364
117411733	0.78	0.86	**0.95**	0.93
117960634	/	/	0.84	0.86
121147495	/	/	0.92	**0.95**

We created a new marker map by 1) keeping the first segment unchanged (all markers till position 117411733), 2) keeping the same marker distances within segments, 3) putting a 292199 bp interval between segment 1 and 3 (the distance between these markers on the Bos_taurus_UMD3.1 assembly, where they are consecutive) and 4) inverting the Baylor Btau_4.6.1 117960634-121147495 segment and locating it 500000 bp from the last position. We rephased the data with this new map and found no evidence of major map errors (Fig. 4).

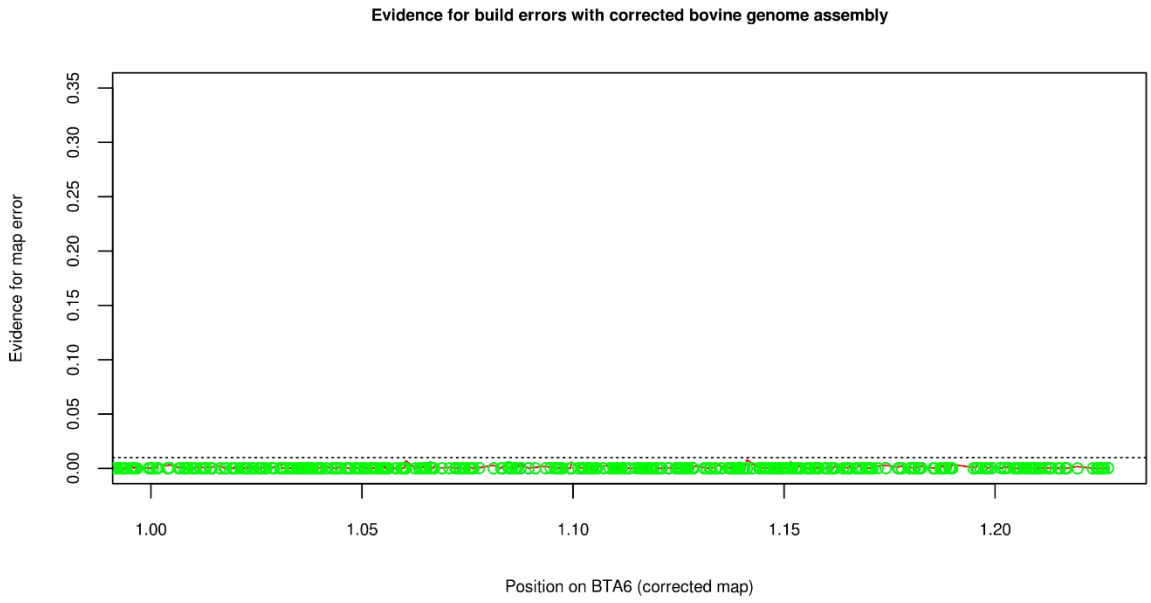


Figure 4. Evidence for errors on BTA6 (corrected marker map) after run of LINKPHASE3: estimated recombination rates in marker intervals (red), parent genotyping error rate (blue), within haplotype allelic entropy (green), overall Map Confidence Score (grey).

To summarize, a 3,186,861 bp long segment (containing 42 markers) has been inverted and moved to the end of the map. Within the segment, the map matches perfectly Baylor Btau_4.6.1 bovine genome assembly.

Supplemental Note S5. Impact of familial information on crossover identification

Introduction. Detection of CO relies on comparison of the parental haplotypes (proband) with transmitted haplotypes (offspring). Therefore, it depends on the accuracy of haplotype reconstruction with LINKPHASE3. In the proband, phasing accuracy is a function of the number of genotyped parents (0, 1 and 2) and the number of genotyped offspring (1 or more). In the offspring, accuracy is improved when the second parent (the mate of the proband) is also genotyped.

Method. To study these effects on CO identification, we selected large paternal half-sibs families (40 or more) with both parents of the proband genotyped. We compared CO identification for five offspring per half-sib family that had also their dam (the mate of the proband) genotyped. In total, comparisons were realized for 255 offspring (we had 51 such half-sib families with at least five offspring with their dam genotyped). The effect of number of genotyped parents of the proband (0, 1 and 2), the number of genotyped offspring (from 1 to 40) and the presence of genotyped mate was assessed by running LINKPHASE3 in subsets of the selected families: by deleting genotyped parents, the mate or by selecting random subsets of genotyped half-sibs.

Results. We compared CO identification for the mean and for the accuracy (measured as the coefficient of determination of a linear regression comparing GRR in a subset of the data with GRR). On average, lower information resulted in fewer detected CO. Both the mean and the accuracy presented asymptotic behavior regarding the number of genotyped offspring. Without parent of the proband genotyped, the plateau was reached with 20 genotyped offspring (although with ten genotyped offspring a high accuracy was already achieved). Only five genotyped offspring were required to reach the plateau in the presence of genotyped parents of the proband. With two genotyped parents, a relatively high accuracy was obtained with a single genotyped offspring. When the

mates are genotyped, ~0.75 additional CO are identified. The mate genotype increases the number of markers phased in the offspring (it does not affect haplotype reconstruction of the proband). Therefore, adding more information to reconstruct the proband haplotype (more genotyped offspring or parents) does not compensate the absence of genotyped mate.

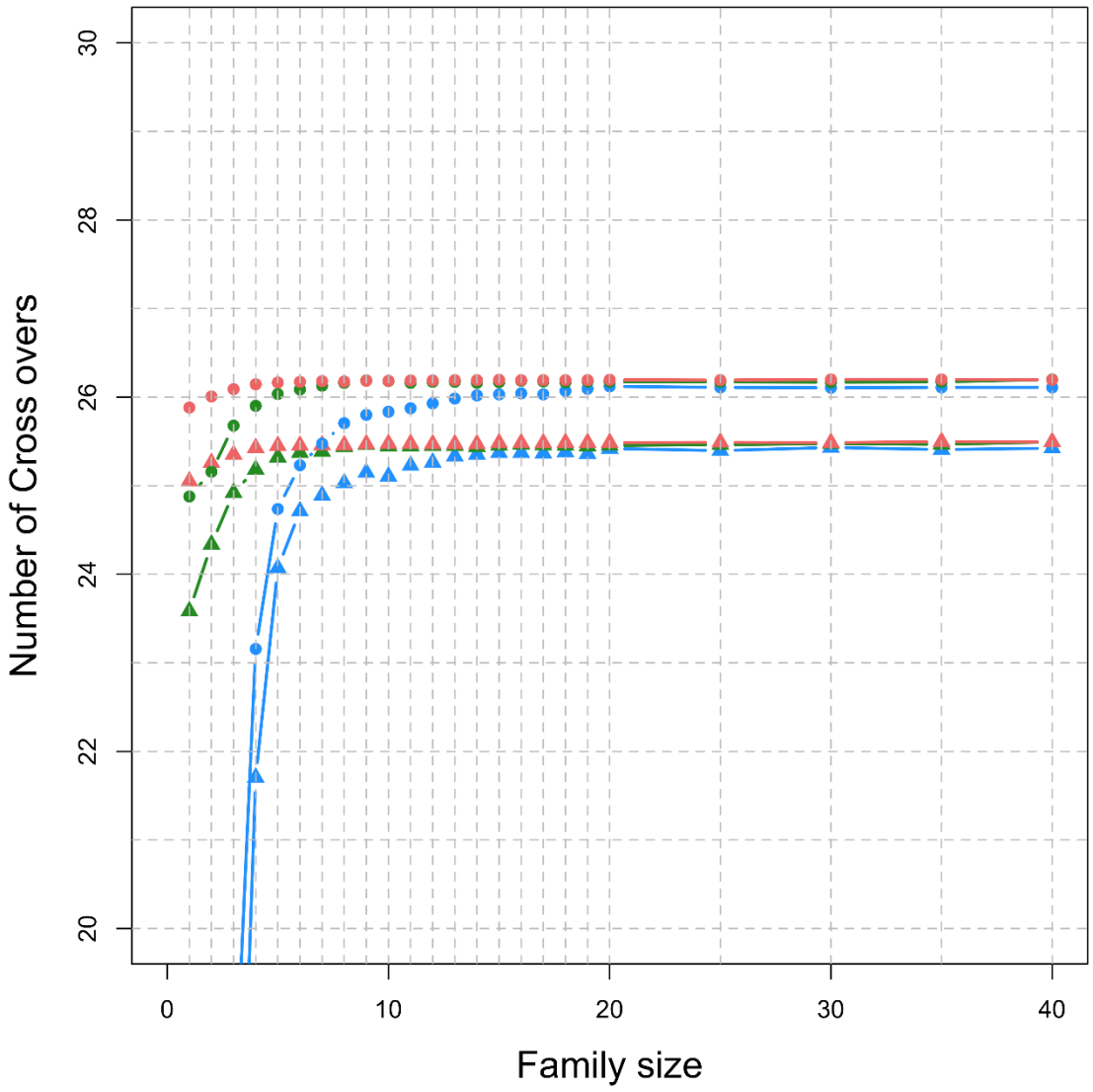


Figure 1. Effect of design on mean CO. Average number of CO identified in half-sib families of varying size (1-40 offsprings; x-axis) and with 0 (blue), 1 (green) or 2 (red) genotyped parents. Circles and triangles indicate whether the mate is genotyped or not.

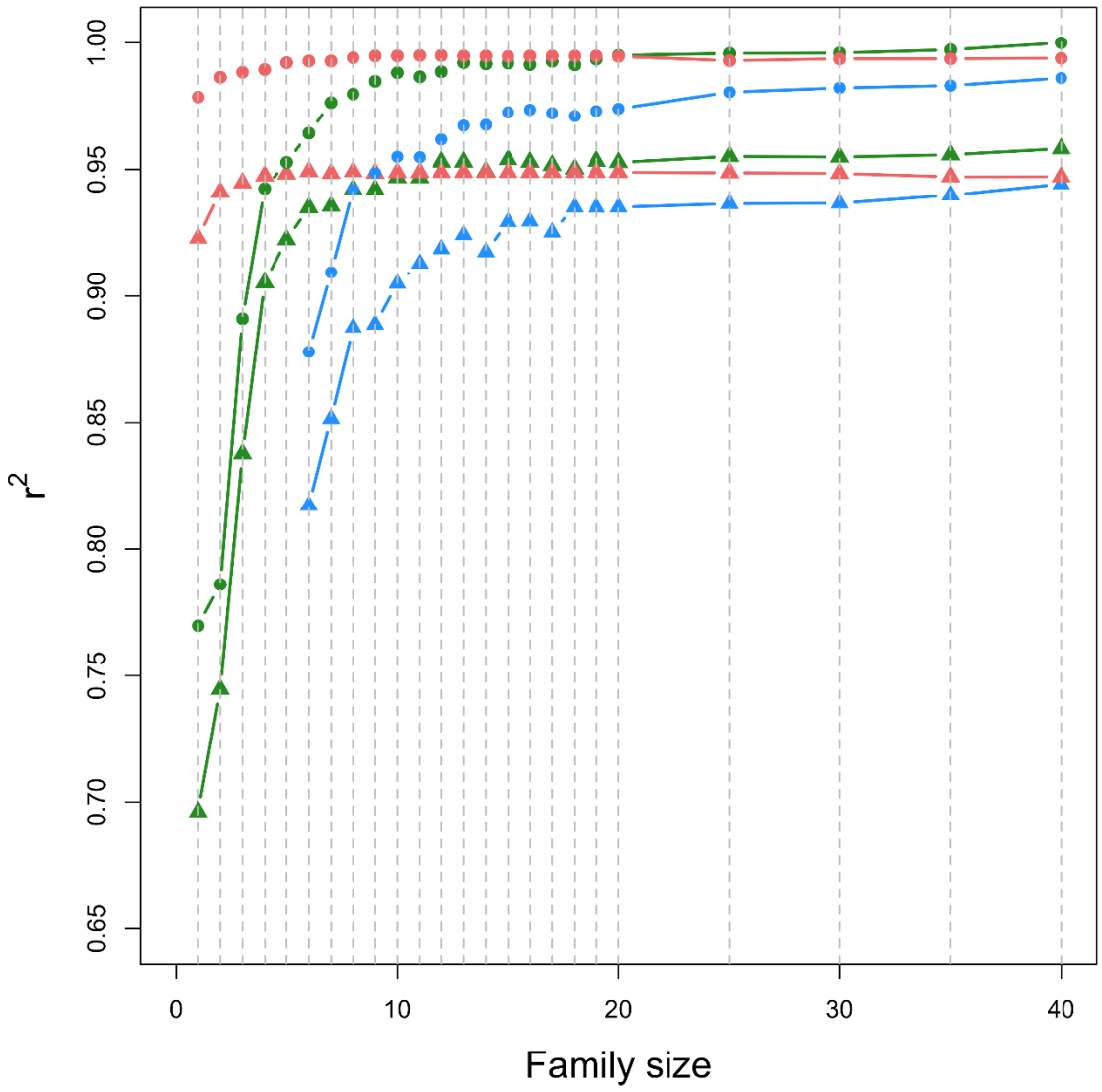


Figure 2. Effect of design on accuracy. Average accuracy of CO identification in half-sib families of varying size (1-40 offsprings; x-axis) and with 0 (blue), 1 (green) or 2 (red) genotyped parents. Circles and triangles indicate whether the mate is genotyped or not.

Conclusions. The amount of familial information has an impact of CO identification. Males and females have different level of familial information (for instance, males have more genotyped offspring). Therefore for comparison of GRR per sex, we will work with standardized designs relying only on genotypes of the offspring, the proband and one of its parents to count GRR associated with the transmission of genetic material from the proband to that offspring. This solution allows to have identical familial information and

to keep as many records as possible. In the raw data, the informativity (measured as the number of heterozygous markers from the proband phased in both the proband and the offspring) is 5,798 in males and 7,527 in females (informativity is larger in females because the mate of the probands are most often genotyped). In the standardized design, identical informativity (4,315 in males and 4,333 in females) is observed (measured on the same individuals as for the raw design).

To maximize power of association studies, as much records as possible were conserved and all the available familial information was used for phasing. We kept all probands with at least one genotyped parent or with more than five genotyped offspring.

Supplemental Note S6. Next Generation Sequencing data and imputation

Next Generation Sequencing samples

Whole-genome sequence from individuals of two cattle populations were used in this study. The first population consisted in 556 individuals from the population from New-Zealand and the second contained 415 individuals of a Dutch Holstein pedigree (called the DAMONA pedigree).

DNA samples were extracted from whole blood or semen using standard protocols. Sequencing was done on Illumina HiSeq 2000 instruments with a PCR free method to prepare libraries with 550bp (DAMONA pedigree) or 350 bp (samples from New-Zealand) insert sizes. Paired-end sequencing with read length of 2 x 100 base pairs was applied.

The whole-genome sequence data was analyzed according to GATK Best Practice V3.4. Alignement of reads (FASTQ files) to the reference genome (Bos Taurus UMD 3.1) was done with BWA MEM (version 0.7.9a-r786, (Li 2013)) with the default settings. The sorted BAM had PCR duplicates detected using sambamba (v0,4,6) and Picard tools and bedtools were used to generate library statistics and coverage information.

The obtained BAM files were then realigned around indels and recalibrated for base quality with Genome Analysis Toolkit (GATK 2.7.4., (DePristo et al. 2011)). List of known SNP used for recalibration were obtained from DBSNP release 138. Variant calling was performed with GATK Haplotype caller in N+1 mode. For calibration of variant quality, a set of trusted SNP and indels was used. For SNPs, the set consisted in SNPs from the BovineHD (Illumina) and Axiom Genome-Wide BOS 1 (Affymetrix) commercial genotyping arrays. For indels, we selected a subset of indels identified in the DAMONA pedigree behaving like true Mendelian variants : presenting no parent-offspring incompatibilities (e.g. opposite homozygotes), no deviation from Hardy-Weinberg

proportions ($p > 0.05$) and no deviation from expected genotypic proportions in offspring of heterozygous parents ($p > 0.05$). In addition, we compute the probability to observe no parent-offspring inconsistency if parental alleles were drawn at random and conserved only indels with a probability below 1e-12 (to make sure that the absence of parent-offspring incompatibilities was not due by chance).

Imputation. To fine-map QTLs identified in the haplotype-based study, variants were imputed in 5 Mb windows around the peak of each QTLs (see Table 1- imputed regions). Imputation was performed within the population from New-Zealand using 122 individuals from the same population sequenced at a cover higher than 15x whereas for the European Holsteins (Dutch and French Holsteins), 215 individuals from the DAMONA Dutch Holstein pedigree sequenced at a cover higher than 15x were selected as reference.

Imputation was realized with Beagle 4.0 (Browning and Browning 2007) and performed in two steps as advised in (van Binsbergen et al. 2014). First, imputation was performed from the map used to study GRR (31,127) to the Illumina BovineHD genotyping array map (with respectively 3539 and 557 genotyped reference individuals in the population from New-Zealand and in French/Dutch Holstein). Next, imputation was performed to the whole-genome sequence level. Association study was performed on imputed variants common to both reference panels after filtering out variants with imputation accuracy below 0.5.

Table 1. Imputed regions, Number of variants used for sequence based association analysis

Imputed regions	50K	New-Zealand		DAMONA		Shared Seq.
		HD	Seq.	HD	Seq.	
chr3:49500000-54500000	47	1,316	33,336	1165	28,308	25,037
chr3 :67000000-72000000	79	1,076	40,334	817	34,235	30,709
chr6:117704799-121164220*	43	598	20,924	533	17,449	15,432
chr10:17500000-23500000	80	1,433	35,419	1095	33,832	28,782
chr10:83500000-88500000	76	1,393	36,557	1151	27,486	24,714
chr18 :50000000-60000000	102	2,348	65,361	1971	59,439	47,735
chr23 :25000000-31000000	51	1,274	125,108	1511	119,545	78,841

*Btau4

For the region on BTA6, we used the Baylor Btau4.6.1 genome assembly. Indeed, the QTL maps at the telomeric end of our new map (described in Supplementary Note 3) corresponding to chr6:117960634-121147495 on Baylor Btau4.6.1 (and poorly assembled on UMD3.1). We reordered accordingly markers from the VCF files (with a LiftOver from UMD3.1 to Baylor Btau4.6.1).

Variant annotation. Annotation of SNPs was performed with snpEFF v4.11 (Cingolani et al. 2012). For *RNF212* region (poorly annotated) and indels, manual annotation was also performed for the variants associated with the most significant variant.

References

Browning SR, Browning BL. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *The American Journal of Human Genetics* **81**: 1084-1097.

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* **6**: 80-92.

Daetwyler HD, Capitan A, Pausch H, Stothard P, Van Binsbergen R, Brøndum RF, Liao X, Djari A, Rodriguez SC, Grohs C. 2014. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nature genetics* **46**: 858-865.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, Del Angel G, Rivas MA, Hanna M. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics* **43**: 491-498.

Druet T, Georges M. 2015. LINKPHASE3: an improved pedigree-based phasing algorithm robust to genotyping and map errors. *Bioinformatics*: btu859.

Druet T, Macleod I, Hayes B. 2014. Toward genomic prediction from whole-genome sequence data: impact of sequencing design on genotype imputation and accuracy of predictions. *Heredity* **112**: 39-47.

Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. 2012. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature genetics* **44**: 955-959.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:13033997*.

van Binsbergen R, Bink M, Calus MP, van Eeuwijk FA, Hayes BJ, Hulsegge I, Veerkamp RF. 2014. Accuracy of imputation to whole-genome sequence data in Holstein Friesian cattle. *Genetics, selection, evolution : GSE* **46**: 41.

Yang J, Lee SH, Goddard ME, Visscher PM. 2011. GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics* **88**: 76-82.